(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 January 2002 (24.01.2002)

PCT

(10) International Publication Number
WO 02/06325 A2

(51) International Patent Classification[7]: C07K 14/435

(21) International Application Number: PCT/US01/22290

(22) International Filing Date: 16 July 2001 (16.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/218,383 14 July 2000 (14.07.2000) US

(71) Applicant *(for all designated States except US)*: **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **MEYERS, Rachel** [US/US]; 115 Devonshire Road, Newton, MA 02468 (US).

(74) Agents: **MANDRAGOURAS, Amy, E.** et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/06325 A2

(54) Title: 32529, A NOVEL HUMAN GUANINE NUCLEOTIDE EXCHANGE FACTOR FAMILY MEMBER AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acids molecules, designated GEF32529 nucleic acid molecules, which encode novel human guanine nucleotide exchange factor (GEF) molecules. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing GEF32529 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a GEF32529 gene has been introduced or disrupted. The invention still further provides isolated GEF32529 polypeptides, fusion polypeptides, antigenic peptides and anti-GEF32529 antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

32529, A NOVEL HUMAN GUANINE NUCLEOTIDE EXCHANGE FACTOR FAMILY MEMBER AND USES THEREOF

Related Applications

This application claims the benefit of prior-filed U.S. Provisional Patent Application No. 60/218,383 filed on July 14, 2000, incorporated herein in its entirety by reference.

Background of the Invention

The G protein superfamily (*e.g.*, heterotrimeric and small G proteins) encompasses a diverse array of proteins which regulate a complex range of biological processes, including the regulation of protein synthesis, cellular trafficking (*e.g.*, vesicular and nuclear transport), regulation of the cell cycle, growth, differentiation, apoptosis, and cytoskeletal rearrangements (Cerione *et al.* (1996) *Curr. Op. Cell Biol.* 8:216-222; Cherfils *et al.* (1999) *Trends Biochem. Sci.* 24:306-311). The common motif among this important family of proteins is the presence of a GTP-binding domain (Alberts *et al.*. (1994) Molecular Biology of the Cell, Garland Publishing, Inc., New York, N.Y. pp. 206-207, 641). These proteins act as molecular switches that can cycle between active (GTP-bound) and inactive (GDP-bound) states (Bourne *et al.* (1990) *Nature*, 348:125-132). In the active state, G proteins are able to interact with a broad range of effector molecules. These effector molecules constitute components of a variety of signaling cascades. Upon hydrolysis of bound GTP, the G protein switches to the inactive state, a step that is facilitated by GTPase activating proteins (GAPs) (Scheffzek *et al.* (1998) *Trends Biochem Sci.* 23:257-262; Gamblin and Smerdon (1998) *Curr. Opinion in Struct. Biol.* 8:195-201).

Activation of G proteins is mediated by the exchange of GDP for GTP. Dissociation of GDP from the inactive small G protein is facilitated by a class of proteins known as guanine nucleotide exchange factors (GEFs). The small G protein is then able to bind GTP and undergo conformational changes which allow it to interact with effector molecules.

GEFs consist of four families based on sequence similarity among family members and on selectivity of small G protein activated by the GEF, including GEFs of Ran, ARF, Ras, and Rho (also known as the Dbl homology (DH) domain-containing GEFs). GEF family members all contain a GEF homology domain amino terminal to a pleckstrin homology (PH) domain, and most contain other functional domains commonly found in signaling molecules (Cerione *et al.* and Cherfils *et al.*, *supra*). For example, the GEF family members Dbs and Vav both have Src homology (SH3) domains at their carboxyl termini (Whitehead *et al.* (1995) *Oncogene* 10:713-721).

Many GEF family members have been identified to date including Dbl, Ost, Tiam-1, Ect-2, Vav, Lbc, FGD1, Dbs, Lfc, Tim, Brc, Abr, Sos, and Ras GEF. These proteins are found in various tissues including adrenal gland, brain, gonad, heart, keratinocyte, kidney, liver, lung, mammary epithelial, myeloid, pancreas, placenta, spleen, skeletal muscle, testis, and fetal brain and heart, and in diffuse B-cell lymphomas, osteosarcomas, T-lymphoma cells, and myeloid leukemias. The Sos protein is ubiquitous (Cerione *et al.*, *supra*).

It is the regulated cycling between active and inactive states of G proteins that allows for proper transduction of many vital cellular signals. Indeed, the regulation of GTP/GDP levels in the cell by small G proteins and their accessory GEF molecules, has been implicated in a number of diseases, including oncogenesis and metastasis, faciogenital dysplasia, chronic myelogenous, and leukemia (Cerione *et al.*, *supra*).

Summary of the Invention

The present invention is based, at least in part, on the discovery of novel guanine nucleotide exchange factor family members, referred to herein as "guanine nucleotide exchange factor-32529" or "GEF32529" nucleic acid and polypeptide molecules. The GEF32529 nucleic acid and polypeptide molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, *e.g.*, cell signaling, tumor inhibition (*e.g.*, growth, differentiation, and apoptosis), cytoskeletal organization (*e.g.*, cell morphology), and cellular trafficking. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding GEF32529 polypeptides or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of GEF32529-encoding nucleic acids.

In one embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence set forth in SEQ ID NO:1 or SEQ ID NO:3. In another embodiment, the invention features an isolated nucleic acid molecule that encodes a polypeptide including the amino acid sequence set forth in SEQ ID NO:2. In another embodiment, the invention features an isolated nucleic acid molecule that includes the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____.

In still other embodiments, the invention features isolated nucleic acid molecules including nucleotide sequences that are substantially identical (*e.g.*, 60% identical) to the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3. The invention further features isolated nucleic acid molecules including at least 30 contiguous nucleotides of the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3. In another embodiment, the invention features isolated nucleic acid molecules which encode a polypeptide including an amino acid sequence that is substantially identical

(*e.g.*, 60% identical) to the amino acid sequence set forth as SEQ ID NO:2. The present invention also features nucleic acid molecules which encode allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:2. In addition to isolated nucleic acid molecules encoding full-length polypeptides, the present invention also features nucleic acid molecules which encode fragments, for example, biologically active or antigenic fragments, of the full-length polypeptides of the present invention (*e.g.*, fragments including at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2). In still other embodiments, the invention features nucleic acid molecules that are complementary to, antisense to, or hybridize under stringent conditions to the isolated nucleic acid molecules described herein.

In a related aspect, the invention provides vectors including the isolated nucleic acid molecules described herein (*e.g.*, GEF32529-encoding nucleic acid molecules). Such vectors can optionally include nucleotide sequences encoding heterologous polypeptides. Also featured are host cells including such vectors (*e.g.*, host cells including vectors suitable for producing GEF32529 nucleic acid molecules and polypeptides).

In another aspect, the invention features isolated GEF32529 polypeptides and/or biologically active or antigenic fragments thereof. Exemplary embodiments feature a polypeptide including the amino acid sequence set forth as SEQ ID NO:2, a polypeptide including an amino acid sequence at least 60% identical to the amino acid sequence set forth as SEQ ID NO:2, a polypeptide encoded by a nucleic acid molecule including a nucleotide sequence at least 60% identical to the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3. Also featured are fragments of the full-length polypeptides described herein (*e.g.*, fragments including at least 10 contiguous amino acid residues of the sequence set forth as SEQ ID NO:2) as well as allelic variants of the polypeptide having the amino acid sequence set forth as SEQ ID NO:2.

The GEF32529 polypeptides and/or biologically active or antigenic fragments thereof, are useful, for example, as reagents or targets in assays applicable to treatment and/or diagnosis of GEF32529 mediated or related disorders. In one embodiment, a GEF32529 polypeptide or fragment thereof, has a GEF32529 activity. In another embodiment, a GEF32529 polypeptide or fragment thereof, has a GEF domain, a signal sequence, and optionally, has a GEF32529 activity. In a related aspect, the invention features antibodies (*e.g.*, antibodies which specifically bind to any one of the polypeptides described herein) as well as fusion polypeptides including all or a fragment of a polypeptide described herein.

The present invention further features methods for detecting GEF32529 polypeptides and/or GEF32529 nucleic acid molecules, such methods featuring, for example, a probe, primer or antibody described herein. Also featured are kits for the

detection of GEF32529 polypeptides and/or GEF32529 nucleic acid molecules. In a related aspect, the invention features methods for identifying compounds which bind to and/or modulate the activity of a GEF32529 polypeptide or GEF32529 nucleic acid molecule described herein. Further featured are methods for modulating a GEF32529 activity.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence and predicted amino acid sequence of human GEF32529. The nucleotide sequence corresponds to nucleic acids 1 to 3075 of SEQ ID NO:1. The amino acid sequence corresponds to amino acids 1 to 802 of SEQ ID NO: 2. The coding region without the 5' and 3' untranslated regions of the human GEF32529 gene is shown in SEQ ID NO: 3.

Figure 2 depicts a structural, hydrophobicity, and antigenicity analysis of the human GEF32529 polypeptide.

Figure 3 depicts the results of a search which was performed against the HMM database in PFAM and SMART and which resulted in the identification of a "GEF domain," a "PH domain," and a "SH3 domain" in the human GEF32529 polypeptide (SEQ ID NO:2).

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as "guanine nucleotide exchange factor-32529" or "GEF32529" nucleic acid and polypeptide molecules, which are novel members of the guanine nucleotide exchange factor family (*e.g.*, the RhoGEF family). These novel molecules are capable of, for example, modulating small G protein mediated activity (*e.g.*, dissociating GDP from a small G protein, for example a Rho/Rac-mediated activity) in a cell, *e.g.*, an adrenal gland, brain, gonad, heart, keratinocyte, kidney, liver, lung, mammary epithelial, myeloid, pancreas, placenta, spleen, skeletal muscle, testis, fetal brain and heart, diffuse B-cell lymphoma, osteosarcoma, T-lymphoma cell, or myeloid leukemia cell. These novel molecules thus, may play a role in or function in a variety of cellular processes, *e.g.*, regulating signal transduction, regulating tumor inhibition, regulating cytoskeletal organization, and/or regulating cellular trafficking. Thus, the GEF32529 molecules of the present invention provide novel diagnostic targets and therapeutic agents to control GEF associated disorders.

As used herein, the term "GEF associated disorder" or "RhoGEF associated disorder" or "Rho/RacGEF associated disorder" includes disorders, diseases, or conditions which are characterized by aberrant, *e.g.*, upregulated or downregulated, GDP dissociation from small G proteins. Examples of such disorders include cancer, inflammation, diabetes, and pathogenic invasion of host cells. Other examples of GEF associated disorders are described herein.

The term "family" when referring to the polypeptide and nucleic acid molecules of the invention is intended to mean two or more polypeptides or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first polypeptide of human origin, as well as other, distinct polypeptides of human origin or alternatively, can contain homologues of non-human origin, *e.g.*, monkey polypeptides. Members of a family may also have common functional characteristics.

In one embodiment, a GEF32529 molecule of the present invention is identified based on the presence of at least one "GEF domain" or "RhoGEF domain" or "Rho/RacGEF domain" As used herein, the term "GEF domain" or "RhoGEF domain" or "Rho/RacGEF domain" includes a protein domain having at least about 80-220 amino acid residues and a bit score of at least 15 when compared against a GEF Hidden Markov Model (HMM in PFAM). Preferably, a GEF domain or RhoGEF domain or Rho/RacGEF domain includes a polypeptide having an amino acid sequence of about 100-200, 110-190, 120-180, or more preferably, about 179 amino acid residues and a bit score of at least 20, 30, 40, 50, or more preferably, 64.5. To identify the presence of a GEF domain in a GEF32529 protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein may be searched against a database of known protein domains (*e.g.*, the PFAM HMM database). A GEF domain HMM (referred to also as RhoGEF) has been assigned the PFAM Accession PF00621 (http://genome.wustl.edu/Pfam/html). A search was performed against the HMM database resulting in the identification of a GEF domain in the amino acid sequence of human GEF32529 (SEQ ID NO:2) at about residues 380-559 of SEQ ID NO:2. The results of the search are set forth in Figure 3.

Preferably a "GEF domain" or "RhoGEF domain" or "Rho/RacGEF domain" has a guanine nucleotide exchange or release activity. Accordingly, identifying the presence of a "GEF domain" or "RhoGEF domain" or "Rho/RacGEF domain" can include isolating a fragment of a GEF32529 molecule (*e.g.*, a GEF32529 polypeptide) and assaying for the ability of the fragment to exchange or release a guanine nucleotide (*e.g.*, GDP) from a guanine nucleotide bound substrate.

In another embodiment, a GEF32529 molecule of the present invention is identified based on the presence of at least one "PH domain." As used herein, the term "PH domain" includes a protein domain having at least about 70-170 amino acid residues and a bit score of at least 10 when compared against a PH Hidden Markov Model (HMM in PFAM). Preferably, a PH domain includes a polypeptide having an amino acid sequence of about 50-150, 60-140, 70-130, 80-120, or more preferably, about 111 amino acid residues and a bit score of at least 15, 20, 25, 30, or more preferably, 33. To identify the presence of a PH domain in a GEF32529 protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein may be searched against a database of known protein domains (*e.g.*, the PFAM HMM database). A PFAM PH domain HMM has been assigned the PFAM Accession PF00169. A search was performed against the PFAM HMM database resulting in the identification of a PH domain in the amino acid sequence of human GEF32529 (SEQ ID NO:2) at about residues 593-704 of SEQ ID NO:2. The results of the search are set forth in Figure 3.

Preferably a "PH domain" has a "PH domain activity," for example, the ability to bind inositol lipids (*e.g.*, phosphatidylinositol lipids), regulate membrane anchoring (*e.g.*, anchoring of the host protein, *i.e.*, the protein containing the domain, to a cellular membrane), modulate enzymatic activity of the host protein (*e.g.*, modulate the activity of adjacent nucleotide exchange domains), target the host protein to a correct subcellular location, and/or respond to upstream signals. Accordingly, identifying the presence of a "PH domain" can include isolating a fragment of a GEF32529 molecule (*e.g.*, a GEF32529 polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned PH domain activities.

In another embodiment, a GEF32529 molecule of the present invention is identified based on the presence of at least one "SH3 domain." As used herein, the term "SH3 domain" includes a protein domain having at least about 5-100 amino acid residues and a bit score of at least 10 when compared against an SH3 Hidden Markov Model (HMM in PFAM). Preferably, an SH3 domain includes a polypeptide having an amino acid sequence of about 10-90, 20-80, 30-70, 40-60, or more preferably, about 50 amino acid residues and a bit score of at least 15, 20, 25, 30, or more preferably, 33. To identify the presence of an SH3 domain in a GEF32529 protein, and make the determination that a protein of interest has a particular profile, the amino acid sequence of the protein may be searched against a database of known protein domains (*e.g.*, the PFAM HMM database). A PFAM SH3 domain HMM has been assigned the PFAM Accession PF00018. A search was performed against the HMM database resulting in the identification of an SH3 domain in the amino acid sequence of human GEF32529

(SEQ ID NO:2) at about residues 724-774 of SEQ ID NO:2. The results of the search are set forth in Figure 3.

Preferably an "SH3 domain" has an "SH3 domain activity," for example, the ability to bind peptides (*e.g.*, proline-rich peptides), regulate signal transduction (*e.g.*, linking signals transmitted from tyrosine kinases at the plasma membrane to effector proteins), and/or modulate cytoskeletal organization (*e.g.*, mediate binding of cytoskeletal proteins to other proteins). Accordingly, identifying the presence of an "SH3 domain" can include isolating a fragment of a GEF32529 molecule (*e.g.*, a GEF32529 polypeptide) and assaying for the ability of the fragment to exhibit one of the aforementioned SH3 domain activities.

A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28:405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.*(1990) *Meth. Enzymol.* 183:146-159; Gribskov *et al.*(1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.*(1994) *J. Mol. Biol.* 235:1501-1531; and Stultz *et al.*(1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference.

In a preferred embodiment, the GEF32529 molecules of the invention include at least one GEF domain, and/or at least one PH domain, and/or at least one SH3 domain.

Isolated polypeptides of the present invention, preferably GEF32529 polypeptides, have an amino acid sequence sufficiently identical to the amino acid sequence of SEQ ID NO:2 or are encoded by a nucleotide sequence sufficiently identical to SEQ ID NO:1 or 3. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (*e.g.*, an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains having at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homology or identity across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as sufficiently identical. Furthermore, amino acid or nucleotide sequences which share at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homology or identity and share a common functional activity are defined herein as sufficiently identical.

In a preferred embodiment, a GEF32529 polypeptide includes at least one or more of the following domains: a GEF domain, a PH domain, and/or an SH3 domain, and has an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%,

85%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous or identical to the amino acid sequence of SEQ ID NO:2, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____'. In yet another preferred embodiment, a GEF32529 polypeptide includes at least one or more of the following domains: a GEF domain, a PH domain, and/or an SH3 domain, and is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3. In another preferred embodiment, a GEF32529 polypeptide includes at least one or more of the following domains: a GEF domain, a PH domain, and/or an SH3 domain, and has a GEF32529 activity.

As used interchangeably herein, a "GEF32529 activity", "biological activity of GEF32529" or "functional activity of GEF32529", refers to an activity exerted by a GEF32529 polypeptide or nucleic acid molecule, for example, in a GEF32529 expressing cell or tissue, or on a GEF32529 target or substrate (*e.g.*, on a GEF32529 binding partner or on a GEF32529 polypeptide, for example, an allosteric activity within the host polypeptide), as determined *in vivo*, or *in vitro*, according to standard techniques. In one embodiment, a GEF32529 activity is a direct activity, such as association with or enzymatic modification of a GEF32529-target molecule. As used herein, a "target molecule" or "binding partner" is a molecule with which a GEF32529 polypeptide binds or interacts in nature, such that GEF32529-mediated function is achieved. A GEF32529 target molecule can be a non-GEF32529 molecule or a GEF32529 polypeptide or polypeptide of the present invention. In an exemplary embodiment, a GEF32529 target molecule is a GEF32529 substrate (*e.g.*, a GEF family domain ligand, for example, GDP-bound to a small G protein). Alternatively, a GEF32529 activity is an indirect activity, such as a cellular signaling activity mediated by interaction of the GEF32529 polypeptide with a GEF32529 substrate or binding partner. The biological activities of GEF32529 are described herein. For example, the GEF32529 polypeptides of the present invention can have one or more of the following activities: (1) association with a GEF32529 substrate or binding partner (*e.g.*, a GDP-bound small G protein, for example, a Ras-like or Rho/Rac-like small G protein); (2) dissociation of GDP from a GEF32529 substrate or binding partner (*e.g.*, a GDP-bound small G protein); (3) destabilization of a GDP-bound small G protein; (4) stabilization of a nucleotide-free small G protein, and (5) activation of a GEF32529 substrate or binding partner; (6) modulation of signal transduction (*e.g.*, signal transduction cascades involving small GTP-binding proteins); (7) control of cell morphology; (8) modulation of adhesion and/or motility of cells; (9) mediation of cytoskeletal organization or

reorganization; (10) modulation of cellular trafficking (*e.g.*, vesicular transport); and (11) modulation of tumor inhibition.

Accordingly, another embodiment of the invention features isolated GEF32529 polypeptides and polypeptides having a GEF32529 activity. Preferred polypeptides are GEF32529 polypeptides having at least one or more of the following domains: a GEF domain, a PH domain, and/or an SH3 domain, and, preferably, a GEF32529 activity.

Additional preferred polypeptides have one or more of the following domains: a GEF domain, a PH domain, and/or an SH3 domain, and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3.

The nucleotide sequence of the isolated human GEF32529 cDNA and the predicted amino acid sequence of the human GEF32529 polypeptide are shown in Figure 1 and in SEQ ID NOs:1 and 2, respectively. A plasmid containing the nucleotide sequence encoding human GEF32529 was deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The human GEF32529 gene, which is approximately 3075 nucleotides in length, encodes a polypeptide which is approximately 802 amino acid residues in length.

Various aspects of the invention are described in further detail in the following subsections:

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode GEF32529 polypeptides or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify GEF32529-encoding nucleic acid molecules (*e.g.*, GEF32529 mRNA) and fragments for use as PCR primers for the amplification or mutation of GEF32529 nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA) and RNA molecules (*e.g.*, mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

The term "isolated nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regard to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated GEF32529 nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, as a hybridization probe, GEF32529 nucleic acid molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to GEF32529 nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

In one embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to the human GEF32529 cDNA. This cDNA comprises sequences encoding the human GEF32529 polypeptide (*i.e.*, "the coding region", from nucleotides 186-2595) as well as 5' untranslated sequences (nucleotides 1-185) and 3' untranslated sequences (nucleotides 2596-3075). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (*e.g.*, nucleotides 186-2595, corresponding to SEQ ID NO:3). Accordingly, in another embodiment, the isolated nucleic acid molecule comprises SEQ ID NO:3 and nucleotides 1-185 and 2596-3075 of SEQ ID NO:1. In yet another embodiment, the nucleic acid molecule consists of the nucleotide sequence set forth as SEQ ID NO:1 or SEQ ID NO:3.

In still another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to the nucleotide sequence shown in SEQ ID NO:1 or 3 (*e.g.*, to the entire length of the nucleotide sequence), or to the nucleotide sequence (*e.g.*, the entire length of the nucleotide sequence) of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or to a portion or complement of any of these nucleotide sequences. In one embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least (or no greater than) 50-100, 100-250, 250-500, 500-750, 750-1000, 1000-1250, 1250-1500, 1500-1700, 1700-1950, 1950-2200, 2200-2450, 2450-2700, 2700-3000 or more nucleotides in length and hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, for example, a fragment which can be used as a probe or primer or a fragment encoding a portion of a GEF32529 polypeptide, *e.g.*, a biologically active portion of a GEF32529 polypeptide. The nucleotide sequence determined from the cloning of the GEF32529 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other GEF32529 family members, as well as GEF32529 homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide. The probe/primer (*e.g.*, oligonucleotide) typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, 75, 80, 85, 90, 95, or 100 or more consecutive nucleotides of a sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, of an anti-sense sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____.

Exemplary probes or primers are at least (or no greater than) 12 or 15, 20 or 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75 or more nucleotides in length and/or comprise consecutive nucleotides of an isolated nucleic acid molecule described herein. Also included within the scope of the present invention are probes or primers comprising contiguous or consecutive nucleotides of an isolated nucleic acid molecule described herein, but for the difference of 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 bases within the probe or primer sequence. Probes based on the GEF32529 nucleotide sequences can be used to detect (*e.g.*, specifically detect) transcripts or genomic sequences encoding the same or homologous polypeptides. In preferred embodiments, the probe further comprises a label group attached thereto, *e.g.*, the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. In another embodiment a set of primers is provided, *e.g.*, primers suitable for use in a PCR, which can be used to amplify a selected region of a GEF32529 sequence, *e.g.*, a domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differs by no greater than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 bases when compared to a sequence disclosed herein or to the sequence of a naturally occurring variant. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a GEF32529 polypeptide, such as by measuring a level of a GEF32529-

encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting GEF32529 mRNA levels or determining whether a genomic GEF32529 gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a GEF32529 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, which encodes a polypeptide having a GEF32529 biological activity (the biological activities of the GEF32529 polypeptides are described herein), expressing the encoded portion of the GEF32529 polypeptide (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the GEF32529 polypeptide. In an exemplary embodiment, the nucleic acid molecule is at least 50-100, 100-250, 250-500, 500-700, 700-1000, 1000-1250, 1250-1500, 1500-1700, 1700-1950, 1950-2200, 2200-2450, 2450-2700, 2700-3000 or more nucleotides in length and encodes a polypeptide having a GEF32529 activity (as described herein).

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. Such differences can be due to due to degeneracy of the genetic code, thus resulting in a nucleic acid which encodes the same GEF32529 polypeptides as those encoded by the nucleotide sequence shown in SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a polypeptide having an amino acid sequence which differs by at least 1, but no greater than 5, 10, 20, 50 or 100 amino acid residues from the amino acid sequence shown in SEQ ID NO:2, or the amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____. In yet another embodiment, the nucleic acid molecule encodes the amino acid sequence of human GEF32529. If an alignment is needed for this comparison, the sequences should be aligned for maximum homology.

Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologues (different locus), and orthologues (different organism) or can be non-naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions (as compared in the encoded product).

Allelic variants result, for example, from DNA sequence polymorphisms within a population (*e.g.*, the human population) that lead to changes in the amino acid sequences of the GEF32529 polypeptides. Such genetic polymorphisms in the GEF32529 genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a GEF32529 polypeptide, preferably a mammalian GEF32529 polypeptide, and can further include non-coding regulatory sequences, and introns.

Accordingly, in one embodiment, the invention features isolated nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with ATCC as Accession Number _____, wherein the nucleic acid molecule hybridizes to a complement of a nucleic acid molecule comprising SEQ ID NO:1 or SEQ ID NO:3, for example, under stringent hybridization conditions.

Allelic variants of human GEF32529 include both functional and non-functional GEF32529 polypeptides. Functional allelic variants are naturally occurring amino acid sequence variants of the human GEF32529 polypeptide that maintain the ability to bind a GEF32529 ligand or substrate and/or modulate GDP dissociation or signal transduction. Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the polypeptide.

Non-functional allelic variants are naturally occurring amino acid sequence variants of the human GEF32529 polypeptide that do not have the ability to mediate nucleoside hydrolysis. Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion or deletion in critical residues or critical regions.

The present invention further provides non-human orthologues (*e.g.*, non-human orthologues of the human GEF32529 polypeptide). Orthologues of the human GEF32529 polypeptides are polypeptides that are isolated from non-human organisms and possess the same GEF32529 ligand binding and/or modulation of membrane excitation mechanisms of the human GEF32529 polypeptide. Orthologues of the human GEF32529 polypeptide can readily be identified as comprising an amino acid sequence that is substantially identical to SEQ ID NO:2.

Moreover, nucleic acid molecules encoding other GEF32529 family members and, thus, which have a nucleotide sequence which differs from the GEF32529 sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the

plasmid deposited with ATCC as Accession Number _____ are intended to be within the scope of the invention. For example, another GEF32529 cDNA can be identified based on the nucleotide sequence of human GEF32529. Moreover, nucleic acid molecules encoding GEF32529 polypeptides from different species, and which, thus, have a nucleotide sequence which differs from the GEF32529 sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ are intended to be within the scope of the invention. For example, a mouse GEF32529 cDNA can be identified based on the nucleotide sequence of a human GEF32529.

Nucleic acid molecules corresponding to natural allelic variants and homologues of the GEF32529 cDNAs of the invention can be isolated based on their homology to the GEF32529 nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. Nucleic acid molecules corresponding to natural allelic variants and homologues of the GEF32529 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the GEF32529 gene.

Orthologues, homologues and allelic variants can be identified using methods known in the art (*e.g.*, by hybridization to an isolated nucleic acid molecule of the present invention, for example, under stringent hybridization conditions). In one embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. In other embodiment, the nucleic acid is at least 100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000, 1000-1050, 1050-1070, 1070-1100, 1100-1150, 1150-1200, 1200-1250, 1250-1300, 1300-1350, 1350-1400, 1400-1450, 1450-1500, 1500-1550, 1550-1600, 1600-1650, 1650-1700, 1700-1950, 1950-2200, 2200-2450, 2450-2700, 2700-3000 or more nucleotides in length.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences that are significantly identical or homologous to each other remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% identical to each other remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., John Wiley & Sons, Inc. (1995), sections 2, 4 and 6. Additional

stringent conditions can be found in *Molecular Cloning: A Laboratory Manual*, Sambrook *et al.*, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989), chapters 7, 9 and 11. A preferred, non-limiting example of stringent hybridization conditions includes hybridization in 4X sodium chloride/sodium citrate (SSC), at about 65-70°C (or hybridization in 4X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 1X SSC, at about 65-70°C. A preferred, non-limiting example of highly stringent hybridization conditions includes hybridization in 1X SSC, at about 65-70°C (or hybridization in 1X SSC plus 50% formamide at about 42-50°C) followed by one or more washes in 0.3X SSC, at about 65-70°C. A preferred, non-limiting example of reduced stringency hybridization conditions includes hybridization in 4X SSC, at about 50-60°C (or alternatively hybridization in 6X SSC plus 50% formamide at about 40-45° C) followed by one or more washes in 2X SSC, at about 50-60°C. Ranges intermediate to the above-recited values, *e.g.*, at 65-70°C or at 42-50°C are also intended to be encompassed by the present invention. SSPE (1xSSPE is 0.15M NaCl, 10mM NaH_2PO_4, and 1.25mM EDTA, pH 7.4) can be substituted for SSC (1xSSC is 0.15M NaCl and 15mM sodium citrate) in the hybridization and wash buffers; washes are performed for 15 minutes each after hybridization is complete. The hybridization temperature for hybrids anticipated to be less than 50 base pairs in length should be 5-10°C less than the melting temperature (T_m) of the hybrid, where T_m is determined according to the following equations. For hybrids less than 18 base pairs in length, $T_m(°C) = 2(\#\text{ of A + T bases}) + 4(\#\text{ of G + C bases})$. For hybrids between 18 and 49 base pairs in length, $T_m(°C) = 81.5 + 16.6(\log_{10}[Na^+]) + 0.41(\%G+C) - (600/N)$, where N is the number of bases in the hybrid, and $[Na^+]$ is the concentration of sodium ions in the hybridization buffer ($[Na^+]$ for 1xSSC = 0.165 M). It will also be recognized by the skilled practitioner that additional reagents may be added to hybridization and/or wash buffers to decrease non-specific hybridization of nucleic acid molecules to membranes, for example, nitrocellulose or nylon membranes, including but not limited to blocking agents (*e.g.*, BSA or salmon or herring sperm carrier DNA), detergents (*e.g.*, SDS), chelating agents (*e.g.*, EDTA), Ficoll, PVP and the like. When using nylon membranes, in particular, an additional preferred, non-limiting example of stringent hybridization conditions is hybridization in 0.25-0.5M NaH_2PO_4, 7% SDS at about 65°C, followed by one or more washes at 0.02M NaH_2PO_4, 1% SDS at 65°C, see *e.g.*, Church and Gilbert (1984) *Proc. Natl. Acad. Sci. USA* 81:1991-1995, (or, alternatively, 0.2X SSC, 1% SDS).

Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1 or 3 and corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring"

nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural polypeptide).

In addition to naturally-occurring allelic variants of the GEF32529 sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, thereby leading to changes in the amino acid sequence of the encoded GEF32529 polypeptides, without altering the functional ability of the GEF32529 polypeptides. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of GEF32529 (*e.g.*, the sequence of SEQ ID NO:2) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are conserved among the GEF32529 polypeptides of the present invention, *e.g.*, those present in a GEF domain, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the GEF32529 polypeptides of the present invention and other members of the GEF32529 family are not likely to be amenable to alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding GEF32529 polypeptides that contain changes in amino acid residues that are not essential for activity. Such GEF32529 polypeptides differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a polypeptide, wherein the polypeptide comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2 (*e.g.*, to the entire length of SEQ ID NO:2).

An isolated nucleic acid molecule encoding a GEF32529 polypeptide identical to the polypeptide of SEQ ID NO:2, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded polypeptide. Mutations can be introduced into SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____ by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues.

A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine, tryptophan), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a GEF32529 polypeptide is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a GEF32529 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for GEF32529 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert of the plasmid deposited with ATCC as Accession Number _____, the encoded polypeptide can be expressed recombinantly and the activity of the polypeptide can be determined.

In a preferred embodiment, a mutant GEF32529 polypeptide can be assayed for the ability to (i) associate with a GEF32529 substrate or binding partner (*e.g.*, a GDP-bound small G protein, for example, a Ras-like or Rho/Rac-like small G protein); (ii) dissociate GDP from a GEF32529 substrate or binding partner (*e.g.*, a GDP-bound small G protein); (iii) destabilize a GDP-bound small G protein; (iv) stabilize a nucleotide-free small G protein, and (v) activate a GEF32529 substrate or binding partner. In another example, a mutant GEF32529 polypeptide can be assayed for the ability to: (1) modulate signal transduction (*e.g.*, signal transduction cascades involving small GTP-binding proteins); (2) control cell morphology; (3) modulate adhesion and/or motility of cells; (4) mediate cytoskeletal organization or reorganization; modulate cellular trafficking (*e.g.*, vesicular transport); and (6) modulate tumor inhibition.

In addition to the nucleic acid molecules encoding GEF32529 polypeptides described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. In an exemplary embodiment, the invention provides an isolated nucleic acid molecule which is antisense to a GEF32529 nucleic acid molecule (*e.g.*, is antisense to the coding strand of a GEF32529 nucleic acid molecule). An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a polypeptide, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire GEF32529 coding

strand, or to only a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding GEF32529. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (*e.g.*, the coding region of human GEF32529 corresponds to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding GEF32529. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding GEF32529 disclosed herein (*e.g.*, SEQ ID NO:3), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of GEF32529 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of GEF32529 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of GEF32529 mRNA (*e.g.*, between the –10 and +10 regions of the start site of a gene nucleotide sequence). An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the

antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a GEF32529 polypeptide to thereby inhibit expression of the polypeptide, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave GEF32529 mRNA transcripts to thereby inhibit translation of GEF32529 mRNA. A ribozyme having specificity for a GEF32529-encoding nucleic acid can be designed based upon the nucleotide sequence of a GEF32529 cDNA disclosed herein (*i.e.*, SEQ ID NO:1 or 3, or the nucleotide sequence of the DNA insert

of the plasmid deposited with ATCC as Accession Number _____). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a GEF32529-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, GEF32529 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, GEF32529 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the GEF32529 (*e.g.*, the GEF32529 promoter and/or enhancers) to form triple helical structures that prevent transcription of the GEF32529 gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the GEF32529 nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *et al.* *Proc. Natl. Acad. Sci.* 93: 14670-675.

PNAs of GEF32529 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of GEF32529 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (*e.g.*, by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (*e.g.*, S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of GEF32529 can be modified, (*e.g.*, to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques

of drug delivery known in the art. For example, PNA-DNA chimeras of GEF32529 nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (*e.g.*, RNase H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. *et al.* (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. *et al.* (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. *et al.* (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. *et al.* (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, *e.g.*, Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

Alternatively, the expression characteristics of an endogenous GEF32529 gene within a cell line or microorganism may be modified by inserting a heterologous DNA regulatory element into the genome of a stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous GEF32529 gene. For example, an endogenous GEF32529 gene which is normally "transcriptionally silent", *i.e.*, a GEF32529 gene which is normally not expressed, or is expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, a transcriptionally silent, endogenous GEF32529 gene may be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with an endogenous GEF32529 gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art, and described, *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

II. Isolated GEF32529 Polypeptides and Anti-GEF32529 Antibodies

One aspect of the invention pertains to isolated GEF32529 or recombinant polypeptides, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-GEF32529 antibodies. In one embodiment, native GEF32529 polypeptides can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, GEF32529 polypeptides are produced by recombinant DNA techniques. Alternative to recombinant expression, a GEF32529 polypeptide or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the GEF32529 polypeptide is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of GEF32529 polypeptide in which the polypeptide is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of GEF32529 polypeptide having less than about 30% (by dry weight) of non-GEF32529 polypeptide (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-GEF32529 polypeptide, still more preferably less than about 10% of non-GEF32529 polypeptide, and most preferably less than about 5% non-GEF32529 polypeptide. When the GEF32529 polypeptide or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of GEF32529 polypeptide in which the polypeptide is separated from chemical precursors or other chemicals which are involved in the synthesis of the polypeptide. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of GEF32529 polypeptide having less than about 30% (by dry weight) of chemical precursors or non-GEF32529 chemicals, more preferably less than about 20% chemical precursors or non-GEF32529

chemicals, still more preferably less than about 10% chemical precursors or non-GEF32529 chemicals, and most preferably less than about 5% chemical precursors or non-GEF32529 chemicals.

As used herein, a "biologically active portion" of a GEF32529 polypeptide includes a fragment of a GEF32529 polypeptide which participates in an interaction between a GEF32529 molecule and a non-GEF32529 molecule. Biologically active portions of a GEF32529 polypeptide include peptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the GEF32529 polypeptide, *e.g.*, the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length GEF32529 polypeptides, and exhibit at least one activity of a GEF32529 polypeptide. Typically, biologically active portions comprise a domain or motif with at least one activity of the GEF32529 polypeptide, *e.g.*, dissociating GDP from a small G protein. A biologically active portion of a GEF32529 polypeptide can be a polypeptide which is, for example, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 525, 550, 575, 600, 625, 650, 675, 700, 725, 750, 775 or 800 or more amino acids in length. Biologically active portions of a GEF32529 polypeptide can be used as targets for developing agents which modulate a GEF32529 mediated activity, *e.g.*, dissociating GDP from a small G protein.

In one embodiment, a biologically active portion of a GEF32529 polypeptide comprises at least one GEF domain. It is to be understood that a preferred biologically active portion of a GEF32529 polypeptide of the present invention comprises at least one or more of the following domains: a GEF domain, a PH domain, and/or an SH3 domain. Moreover, other biologically active portions, in which other regions of the polypeptide are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native GEF32529 polypeptide.

Another aspect of the invention features fragments of the polypeptide having the amino acid sequence of SEQ ID NO:2, for example, for use as immunogens. In one embodiment, a fragment comprises at least 5 amino acids (*e.g.*, contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____. In another embodiment, a fragment comprises at least 10, 15, 20, 25, 30, 35, 40, 45, 50 or more amino acids (*e.g.*, contiguous or consecutive amino acids) of the amino acid sequence of SEQ ID NO:2, or an amino acid sequence encoded by the DNA insert of the plasmid deposited with the ATCC as Accession Number _____.

In a preferred embodiment, a GEF32529 polypeptide has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the GEF32529 polypeptide is substantially identical to SEQ ID NO:2, and retains the functional activity of the polypeptide of SEQ ID NO:2, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. In another embodiment, the GEF32529 polypeptide is a polypeptide which comprises an amino acid sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO:2.

In another embodiment, the invention features a GEF32529 polypeptide which is encoded by a nucleic acid molecule consisting of a nucleotide sequence at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more identical to a nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, or a complement thereof. This invention further features a GEF32529 polypeptide which is encoded by a nucleic acid molecule consisting of a nucleotide sequence which hybridizes under stringent hybridization conditions to a complement of a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, or a complement thereof.

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the GEF32529 amino acid sequence of SEQ ID NO:2 having 802 amino acid residues, at least 241, preferably at least 321, more preferably at least 401, more preferably at least 481, even more preferably at least 561, and even more preferably at least 642 or 722 or more amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A preferred, non-limiting example of parameters to be used in conjunction with the GAP program include a Blosum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (*Comput. Appl. Biosci.*, 4:11-17 (1988)) which has been incorporated into the ALIGN program (version 2.0 or version 2.0U), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and polypeptide sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to GEF32529 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 100, wordlength = 3, and a Blosum62 matrix to obtain amino acid sequences homologous to GEF32529 polypeptide molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

The invention also provides GEF32529 chimeric or fusion proteins. As used herein, a GEF32529 "chimeric protein" or "fusion protein" comprises a GEF32529 polypeptide operatively linked to a non-GEF32529 polypeptide. A "GEF32529 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to GEF32529, whereas a "non-GEF32529 polypeptide" refers to a polypeptide having an

amino acid sequence corresponding to a polypeptide which is not substantially homologous to the GEF32529 polypeptide, *e.g.*, a polypeptide which is different from the GEF32529 polypeptide and which is derived from the same or a different organism. Within a GEF32529 fusion protein the GEF32529 polypeptide can correspond to all or a portion of a GEF32529 polypeptide. In a preferred embodiment, a GEF32529 fusion protein comprises at least one biologically active portion of a GEF32529 polypeptide. In another preferred embodiment, a GEF32529 fusion protein comprises at least two biologically active portions of a GEF32529 polypeptide. Within the fusion protein, the term "operatively linked" is intended to indicate that the GEF32529 polypeptide and the non-GEF32529 polypeptide are fused in-frame to each other. The non-GEF32529 polypeptide can be fused to the N-terminus or C-terminus of the GEF32529 polypeptide.

For example, in one embodiment, the fusion protein is a GST-GEF32529 fusion protein in which the GEF32529 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant GEF32529. In another embodiment, the fusion protein is a GEF32529 polypeptide containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of GEF32529 can be increased through the use of a heterologous signal sequence.

The GEF32529 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The GEF32529 fusion proteins can be used to affect the bioavailability of a GEF32529 substrate. Use of GEF32529 fusion proteins may be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a GEF32529 polypeptide; (ii) mis-regulation of the GEF32529 gene; and (iii) aberrant post-translational modification of a GEF32529 polypeptide.

Moreover, the GEF32529-fusion proteins of the invention can be used as immunogens to produce anti-GEF32529 antibodies in a subject, to purify GEF32529 ligands and in screening assays to identify molecules which inhibit the interaction of GEF32529 with a GEF32529 substrate.

Preferably, a GEF32529 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor

primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A GEF32529-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the GEF32529 polypeptide.

The present invention also pertains to variants of the GEF32529 polypeptides which function as either GEF32529 agonists (mimetics) or as GEF32529 antagonists. Variants of the GEF32529 polypeptides can be generated by mutagenesis, *e.g.*, discrete point mutation or truncation of a GEF32529 polypeptide. An agonist of the GEF32529 polypeptides can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a GEF32529 polypeptide. An antagonist of a GEF32529 polypeptide can inhibit one or more of the activities of the naturally occurring form of the GEF32529 polypeptide by, for example, competitively modulating a GEF32529-mediated activity of a GEF32529 polypeptide. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the polypeptide has fewer side effects in a subject relative to treatment with the naturally occurring form of the GEF32529 polypeptide.

In one embodiment, variants of a GEF32529 polypeptide which function as either GEF32529 agonists (mimetics) or as GEF32529 antagonists can be identified by screening combinatorial libraries of mutants, *e.g.*, truncation mutants, of a GEF32529 polypeptide for GEF32529 polypeptide agonist or antagonist activity. In one embodiment, a variegated library of GEF32529 variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of GEF32529 variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential GEF32529 sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of GEF32529 sequences therein. There are a variety of methods which can be used to produce libraries of potential GEF32529 variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential GEF32529 sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, *e.g.*, Narang, S.A. (1983) *Tetrahedron* 39:3;

Itakura *et al.* (1984) *Annu. Rev. Biochem.* 53:323; Itakura *et al.* (1984) *Science* 198:1056; Ike *et al.* (1983) *Nucleic Acid Res.* 11:477.

In addition, libraries of fragments of a GEF32529 polypeptide coding sequence can be used to generate a variegated population of GEF32529 fragments for screening and subsequent selection of variants of a GEF32529 polypeptide. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a GEF32529 coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the GEF32529 polypeptide.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of GEF32529 polypeptides. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify GEF32529 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated GEF32529 library. For example, a library of expression vectors can be transfected into a cell line, *e.g.*, an endothelial cell line, which ordinarily responds to GEF32529 in a particular GEF32529 substrate-dependent manner. The transfected cells are then contacted with GEF32529 and the effect of expression of the mutant on signaling by the GEF32529 substrate can be detected, *e.g.*, by monitoring intracellular GDP concentrations. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the GEF32529 substrate, and the individual clones further characterized.

An isolated GEF32529 polypeptide, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind GEF32529 using standard techniques for polyclonal and monoclonal antibody preparation. A full-length GEF32529 polypeptide can be used or, alternatively, the invention provides antigenic peptide fragments of GEF32529 for use as immunogens. The antigenic peptide of GEF32529 comprises at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of GEF32529 such that an antibody raised against the peptide forms a specific immune complex with GEF32529. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

Preferred epitopes encompassed by the antigenic peptide are regions of GEF32529 that are located on the surface of the polypeptide, *e.g.*, hydrophilic regions, as well as regions with high antigenicity (see, for example, Figure 2).

A GEF32529 immunogen typically is used to prepare antibodies by immunizing a suitable subject, (*e.g.*, rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed GEF32529 polypeptide or a chemically synthesized GEF32529 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic GEF32529 preparation induces a polyclonal anti-GEF32529 antibody response.

Accordingly, another aspect of the invention pertains to anti-GEF32529 antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as GEF32529. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind GEF32529. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of GEF32529. A monoclonal antibody composition thus typically displays a single binding affinity for a particular GEF32529 polypeptide with which it immunoreacts.

Polyclonal anti-GEF32529 antibodies can be prepared as described above by immunizing a suitable subject with a GEF32529 immunogen. The anti-GEF32529 antibody titer in the immunized subject can be monitored over time by standard

techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized GEF32529. If desired, the antibody molecules directed against GEF32529 can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the anti-GEF32529 antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown *et al.* (1981) *J. Immunol.* 127:539-46; Brown *et al.* (1980) *J. Biol. Chem* .255:4980-83; Yeh *et al.* (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh *et al.* (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor *et al.* (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole *et al.* (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter *et al.* (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a GEF32529 immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds GEF32529.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-GEF32529 monoclonal antibody (see, *e.g.*, G. Galfre *et al.* (1977) *Nature* 266:55052; Gefter *et al.* *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful. Typically, the immortal cell line (*e.g.*, a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine, aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, *e.g.*, the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes using polyethylene glycol ("PEG"). Hybridoma

cells resulting from the fusion are then selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind GEF32529, *e.g.*, using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-GEF32529 antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with GEF32529 to thereby isolate immunoglobulin library members that bind GEF32529. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner *et al.* U.S. Patent No. 5,223,409; Kang *et al.* PCT International Publication No. WO 92/18619; Dower *et al.* PCT International Publication No. WO 91/17271; Winter *et al.* PCT International Publication WO 92/20791; Markland *et al.* PCT International Publication No. WO 92/15679; Breitling *et al.* PCT International Publication WO 93/01288; McCafferty *et al.* PCT International Publication No. WO 92/01047; Garrard *et al.* PCT International Publication No. WO 92/09690; Ladner *et al.* PCT International Publication No. WO 90/02809; Fuchs *et al.* (1991) *Bio/Technology* 9:1370-1372; Hay *et al.* (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse *et al.* (1989) *Science* 246:1275-1281; Griffiths *et al.* (1993) *EMBO J* 12:725-734; Hawkins *et al.* (1992) *J. Mol. Biol.* 226:889-896; Clarkson *et al.* (1991) *Nature* 352:624-628; Gram *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrad *et al.* (1991) *Bio/Technology* 9:1373-1377; Hoogenboom *et al.* (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty *et al. Nature* (1990) 348:552-554.

Additionally, recombinant anti-GEF32529 antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.*

(1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison, S. L. (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

An anti-GEF32529 antibody (*e.g.*, monoclonal antibody) can be used to isolate GEF32529 by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-GEF32529 antibody can facilitate the purification of natural GEF32529 from cells and of recombinantly produced GEF32529 expressed in host cells. Moreover, an anti-GEF32529 antibody can be used to detect GEF32529 polypeptide (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the GEF32529 polypeptide. Anti-GEF32529 antibodies can be used diagnostically to monitor polypeptide levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, for example recombinant expression vectors, containing a nucleic acid containing a GEF32529 nucleic acid molecule or vectors containing a nucleic acid molecule which encodes a GEF32529 polypeptide (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are

introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, GEF32529 polypeptides, mutant forms of GEF32529 polypeptides, fusion proteins, and the like).

Accordingly, an exemplary embodiment provides a method for producing a polypeptide, preferably a GEF32529 polypeptide, by culturing in a suitable medium a host cell of the invention (*e.g.*, a mammalian host cell such as a non-human mammalian cell) containing a recombinant expression vector, such that the polypeptide is produced.

The recombinant expression vectors of the invention can be designed for expression of GEF32529 polypeptides in prokaryotic or eukaryotic cells. For example, GEF32529 polypeptides can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in GEF32529 activity assays, (*e.g.*, direct assays or competitive assays described in detail below), or to generate antibodies specific for GEF32529 polypeptides, for example. In a preferred embodiment, a GEF32529 fusion protein expressed in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (*e.g.*, six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident

prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the GEF32529 expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz *et al.*, (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (InVitrogen Corp, San Diego, CA).

Alternatively, GEF32529 polypeptides can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.* (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman *et al.* (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988)

Adv. Immunol. 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to GEF32529 mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. *et al.*, Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a GEF32529 nucleic acid molecule of the invention is introduced, *e.g.*, a GEF32529 nucleic acid molecule within a vector (*e.g.*, a recombinant expression vector) or a GEF32529 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a GEF32529 polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a GEF32529 polypeptide or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a GEF32529 polypeptide. Accordingly, the invention further provides methods for producing a GEF32529 polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of the invention (into which a recombinant expression vector encoding a GEF32529 polypeptide has been introduced) in a suitable medium such that a GEF32529 polypeptide is produced. In another embodiment, the method further comprises isolating a GEF32529 polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which GEF32529-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous GEF32529 sequences have been introduced into their genome or homologous recombinant animals in which endogenous GEF32529 sequences have been

altered. Such animals are useful for studying the function and/or activity of a GEF32529 and for identifying and/or evaluating modulators of GEF32529 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous GEF32529 gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a GEF32529-encoding nucleic acid into the male pronuclei of a fertilized oocyte, *e.g.*, by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The GEF32529 cDNA sequence of SEQ ID NO:1 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human GEF32529 gene, such as a mouse or rat GEF32529 gene, can be used as a transgene. Alternatively, a GEF32529 gene homologue, such as another GEF32529 family member, can be isolated based on hybridization to the GEF32529 cDNA sequences of SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____ (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a GEF32529 transgene to direct expression of a GEF32529 polypeptide to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a GEF32529 transgene in its genome and/or expression of GEF32529 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the

transgene. Moreover, transgenic animals carrying a transgene encoding a GEF32529 polypeptide can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a GEF32529 gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the GEF32529 gene. The GEF32529 gene can be a human gene (*e.g.*, the cDNA of SEQ ID NO:3), but more preferably, is a non-human homologue of a human GEF32529 gene (*e.g.*, a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1). For example, a mouse GEF32529 gene can be used to construct a homologous recombination nucleic acid molecule, *e.g.*, a vector, suitable for altering an endogenous GEF32529 gene in the mouse genome. In a preferred embodiment, the homologous recombination nucleic acid molecule is designed such that, upon homologous recombination, the endogenous GEF32529 gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the homologous recombination nucleic acid molecule can be designed such that, upon homologous recombination, the endogenous GEF32529 gene is mutated or otherwise altered but still encodes functional polypeptide (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous GEF32529 polypeptide). In the homologous recombination nucleic acid molecule, the altered portion of the GEF32529 gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the GEF32529 gene to allow for homologous recombination to occur between the exogenous GEF32529 gene carried by the homologous recombination nucleic acid molecule and an endogenous GEF32529 gene in a cell, *e.g.*, an embryonic stem cell. The additional flanking GEF32529 nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the homologous recombination nucleic acid molecule (see, *e.g.*, Thomas, K. R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The homologous recombination nucleic acid molecule is introduced into a cell, *e.g.*, an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced GEF32529 gene has homologously recombined with the endogenous GEF32529 gene are selected (see *e.g.*, Li, E. *et al.* (1992) *Cell* 69:915). The selected cells can then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras (see *e.g.*, Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the

homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination nucleic acid molecules, *e.g.*, vectors, or homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec *et al.*; WO 91/01140 by Smithies *et al.*; WO 92/0968 by Zijlstra *et al.*; and WO 93/04169 by Berns *et al.*

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, *e.g.*, Lakso *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman *et al.* (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al.* (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, *e.g.*, a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, *e.g.*, the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The GEF32529 nucleic acid molecules, fragments of GEF32529 polypeptides, and anti-GEF32529 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, polypeptide, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration.

The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, a fragment of a GEF32529 polypeptide or an anti-GEF32529 antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as

determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

As defined herein, a therapeutically effective amount of polypeptide (*i.e.*, an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The skilled artisan will appreciate that certain factors may influence the dosage required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a polypeptide or antibody can include a single treatment or, preferably, can include a series of treatments.

In a preferred example, a subject is treated with antibody or polypeptide in the range of between about 0.1 to 20 mg/kg body weight, one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. It will also be appreciated that the effective dosage of antibody or polypeptide used for treatment may increase or decrease over the course of a particular treatment. Changes in dosage may result and become apparent from the results of diagnostic assays as described herein.

The present invention encompasses agents which modulate expression or activity. An agent may, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (*i.e.*,. including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds. It is understood that appropriate doses of small molecule agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of the small molecule will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the small molecule to have upon the nucleic acid or polypeptide of the invention.

Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (*e.g.*, about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. Such appropriate doses may be determined using the assays described herein. When one or more of these small molecules is to be administered to an animal (*e.g.*, a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher may, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

Further, an antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologues thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin

such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon *et al.*, "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld *et al.* (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom *et al.*, "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson *et al.* (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera *et al.* (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin *et al.* (eds.), pp. 303-16 (Academic Press 1985), and Thorpe *et al.*, "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982). Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, protein fragments, GEF32529 modulators, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c)

methods of treatment (*e.g.*, therapeutic and prophylactic). As described herein, a GEF32529 polypeptide of the invention has one or more of the following activities: (i) association with a GEF32529 substrate or binding partner (*e.g.*, a GDP-bound small G protein, for example, a Ras-like or Rho/Rac-like small G protein); (ii) dissociation of GDP from a GEF32529 substrate or binding partner (*e.g.*, a GDP-bound small G protein); (iii) destabilization of a GDP-bound small G protein; (iv) stabilization of a nucleotide-free small G protein, and (v) activation of a GEF32529 substrate or binding partner. In another example, a GEF32529 activity is at least one or more of the following activities: (1) modulation of signal transduction (*e.g.*, signal transduction cascades involving small GTP-binding proteins); (2) control of cell morphology; (3) modulation of adhesion and/or motility of cells; (4) mediation of cytoskeletal organization or reorganization; (5) modulation of cellular trafficking (*e.g.*, vesicular transport); and (6) modulation of tumor inhibition.

The isolated nucleic acid molecules of the invention can be used, for example, to express GEF32529 polypeptide (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect GEF32529 mRNA (*e.g.*, in a biological sample) or a genetic alteration in a GEF32529 gene, and to modulate GEF32529 activity, as described further below. The GEF32529 polypeptides can be used to treat disorders characterized by insufficient or excessive production of a GEF32529 substrate or production of GEF32529 inhibitors (*i.e.*, a "GEF32529 associated," "GEF associated" or "Rho/Rac GEF associated" disorder). As used herein, the term "GEF associated disorder" includes disorders, diseases, or conditions which are characterized by aberrant, *e.g.*, upregulated or downregulated, GDP dissociation from small G proteins. Examples of such disorders include cancer, inflammation, diabetes, and pathogenic invasion of host cells. Other examples are cardiovascular disorders, *e.g.*, arteriosclerosis, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, atrial fibrillation, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, dilated cardiomyopathy, idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, or arrhythmia.

In another example, the activity of a GEF32529 molecule of the present invention is an oncogenic or metastatic activity. As such, GEF32529 molecules are particularly useful in screening for modulators of oncogenesis and/or metastasis, the modulators further being useful in the prophylactic and/or therapeutic methods described herein.

Other examples of GEF associated disorders include disorders of the central nervous system, *e.g.*, cystic fibrosis, type 1 neurofibromatosis, cognitive and neurodegenerative disorders, examples of which include, but are not limited to, Alzheimer's disease, dementias related to Alzheimer's disease (such as Pick's disease), Parkinson's and other Lewy diffuse body diseases, senile dementia, Huntington's disease, Gilles de la Tourette's syndrome, multiple sclerosis, amyotrophic lateral sclerosis, progressive supranuclear palsy, epilepsy, and Creutzfeldt-Jakob disease; autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders, such as depression, schizophrenia, schizoaffective disorder, korsakoff's psychosis, mania, anxiety disorders, or phobic disorders; learning or memory disorders, *e.g.*, amnesia or age-related memory loss, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, obsessive-compulsive disorder, psychoactive substance use disorders, anxiety, phobias, panic disorder, as well as bipolar affective disorder, *e.g.*, severe bipolar affective (mood) disorder (BP-1), and bipolar affective neurological disorders, *e.g.*, migraine and obesity. Further GEF-related disorders include, for example, those listed in the American Psychiatric Association's Diagnostic and Statistical manual of Mental Disorders (DSM), the most current version of which is incorporated herein by reference in its entirety.

Still other examples of GEF associated disorders include cellular proliferation, growth, differentiation, or migration disorders. Cellular proliferation, growth, differentiation, or migration disorders include those disorders that affect cell proliferation, growth, differentiation, or migration processes. As used herein, a "cellular proliferation, growth, differentiation, or migration process" is a process by which a cell increases in number, size or content, by which a cell develops a specialized set of characteristics which differ from that of other cells, or by which a cell moves closer to or further from a particular location or stimulus. Such disorders include cancer, *e.g.*, carcinoma, sarcoma, or leukemia; tumor angiogenesis and metastasis; skeletal dysplasia; hepatic disorders; and hematopoietic and/or myeloproliferative disorders.

Still other examples of GEF associated disorders include disorders of the immune system, such as Wiskott-Aldrich syndrome, viral infection, autoimmune disorders or immune deficiency disorders, *e.g.*, congenital X-linked infantile hypogammaglobulinemia, transient hypogammaglobulinemia, common variable immunodeficiency, selective IgA deficiency, chronic mucocutaneous candidiasis, or severe combined immunodeficiency. Other examples of GEF-related disorders include congenital malformalities, including facio-genital dysplasia; and skin disorders, including microphthalmia with linear skin defects syndrome.

In addition, the GEF32529 polypeptides can be used to screen for naturally occurring GEF32529 substrates, to screen for drugs or compounds which modulate GEF32529 activity, as well as to treat disorders characterized by insufficient or excessive production of GEF32529 polypeptide or production of GEF32529 polypeptide forms which have decreased, aberrant or unwanted activity compared to GEF32529 wild type polypeptide (*e.g.*, nucleoside hydrolysis disorders (such as cell permeabilization, cell necrosis or apoptosis, triggering of second messengers, cell proliferation, cell motility, or signal transduction disorders)). Moreover, the anti-GEF32529 antibodies of the invention can be used to detect and isolate GEF32529 polypeptides, to regulate the bioavailability of GEF32529 polypeptides, and modulate GEF32529 activity.

A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) which bind to GEF32529 polypeptides, have a stimulatory or inhibitory effect on, for example, GEF32529 expression or GEF32529 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of GEF32529 substrate.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a GEF32529 polypeptide or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a GEF32529 polypeptide or polypeptide or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.* 37:2678; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a GEF32529 polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate GEF32529 activity is determined. Determining the ability of the test compound to modulate GEF32529 activity can be accomplished by monitoring, for example, intracellular GDP concentrations. The cell, for example, can be of mammalian origin, *e.g.*, a heart, placenta, lung, liver, skeletal muscle, thymus, kidney, pancreas, testis, ovary, prostate, colon, or brain cell..

The ability of the test compound to modulate GEF32529 binding to a substrate or to bind to GEF32529 can also be determined. Determining the ability of the test compound to modulate GEF32529 binding to a substrate can be accomplished, for example, by coupling the GEF32529 substrate with a radioisotope or enzymatic label such that binding of the GEF32529 substrate to GEF32529 can be determined by detecting the labeled GEF32529 substrate in a complex. Alternatively, GEF32529 could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate GEF32529 binding to a GEF32529 substrate in a complex. Determining the ability of the test compound to bind GEF32529 can be accomplished, for example, by coupling the compound with a radioisotope or enzymatic label such that binding of the compound to GEF32529 can be determined by detecting the labeled GEF32529 compound in a complex. For example, compounds (*e.g.*, GEF32529 substrates) can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound (*e.g.*, a GEF32529 substrate) to interact with GEF32529 without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a compound with GEF32529 without the labeling of either the compound or the GEF32529. McConnell, H. M. *et al.* (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (*e.g.*, Cytosensor) is an analytical instrument that

measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and GEF32529.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a GEF32529 target molecule (*e.g.*, a GEF32529 substrate) with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the GEF32529 target molecule. Determining the ability of the test compound to modulate the activity of a GEF32529 target molecule can be accomplished, for example, by determining the ability of the GEF32529 polypeptide to bind to or interact with the GEF32529 target molecule.

Determining the ability of the GEF32529 polypeptide, or a biologically active fragment thereof, to bind to or interact with a GEF32529 target molecule can be accomplished by one of the methods described above for determining direct binding. In a preferred embodiment, determining the ability of the GEF32529 polypeptide to bind to or interact with a GEF32529 target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.*, intracellular Ca^{2+}, diacylglycerol, IP_3, and the like), detecting catalytic/enzymatic activity of the target using an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a target-regulated cellular response.

In yet another embodiment, an assay of the present invention is a cell-free assay in which a GEF32529 polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the GEF32529 polypeptide or biologically active portion thereof is determined. Preferred biologically active portions of the GEF32529 polypeptides to be used in assays of the present invention include fragments which participate in interactions with non-GEF32529 molecules, *e.g.*, fragments with high surface probability scores (see, for example, Figure 2). Binding of the test compound to the GEF32529 polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the GEF32529 polypeptide or biologically active portion thereof with a known compound which binds GEF32529 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a GEF32529 polypeptide, wherein determining the ability of the test compound to interact with a GEF32529 polypeptide comprises determining the ability of the test compound to preferentially bind to GEF32529 or biologically active portion thereof as compared to the known compound.

In another embodiment, the assay is a cell-free assay in which a GEF32529 polypeptide or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the GEF32529 polypeptide or biologically active portion thereof is determined. Determining the ability of the test compound to modulate the activity of a GEF32529 polypeptide can be accomplished, for example, by determining the ability of the GEF32529 polypeptide to bind to a GEF32529 target molecule by one of the methods described above for determining direct binding. Determining the ability of the GEF32529 polypeptide to bind to a GEF32529 target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (*e.g.*, BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a GEF32529 polypeptide can be accomplished by determining the ability of the GEF32529 polypeptide to further modulate the activity of a downstream effector of a GEF32529 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a GEF32529 polypeptide or biologically active portion thereof with a known compound which binds the GEF32529 polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the GEF32529 polypeptide, wherein determining the ability of the test compound to interact with the GEF32529 polypeptide comprises determining the ability of the GEF32529 polypeptide to preferentially bind to or modulate the activity of a GEF32529 target molecule.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either GEF32529 or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a GEF32529 polypeptide, or interaction of a GEF32529 polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or

both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ GEF32529 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized micrometer plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or GEF32529 polypeptide, and the mixture incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or micrometer plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of GEF32529 binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a GEF32529 polypeptide or a GEF32529 target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated GEF32529 polypeptide or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with GEF32529 polypeptide or target molecules but which do not interfere with binding of the GEF32529 polypeptide to its target molecule can be derivatized to the wells of the plate, and unbound target or GEF32529 polypeptide trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the GEF32529 polypeptide or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the GEF32529 polypeptide or target molecule.

In another embodiment, modulators of GEF32529 expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of GEF32529 mRNA or polypeptide in the cell is determined. The level of expression of GEF32529 mRNA or polypeptide in the presence of the candidate compound is compared to the level of expression of GEF32529 mRNA or polypeptide in the absence of the candidate compound. The candidate compound can then be identified as a modulator of GEF32529 expression based on this comparison. For example, when expression of GEF32529 mRNA or polypeptide is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of GEF32529 mRNA or polypeptide expression. Alternatively, when expression of GEF32529 mRNA or polypeptide is less (statistically

significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of GEF32529 mRNA or polypeptide expression. The level of GEF32529 mRNA or polypeptide expression in the cells can be determined by methods described herein for detecting GEF32529 mRNA or polypeptide.

In yet another aspect of the invention, the GEF32529 polypeptides can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, *e.g.*, U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with GEF32529 ("GEF32529-binding proteins" or "GEF32529-bp") and are involved in GEF32529 activity. Such GEF32529-binding proteins are also likely to be involved in the propagation of signals by the GEF32529 polypeptides or GEF32529 targets as, for example, downstream elements of a GEF32529-mediated signaling pathway. Alternatively, such GEF32529-binding proteins are likely to be GEF32529 inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a GEF32529 polypeptide is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a GEF32529-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the GEF32529 polypeptide.

In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a GEF32529 polypeptide can be confirmed *in vivo*, *e.g.*, in an animal such as an animal model for cellular transformation and/or tumorigenesis.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a GEF32529 modulating agent, an antisense GEF32529 nucleic acid molecule, a GEF32529-specific antibody, or a GEF32529-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the GEF32529 nucleotide sequences, described herein, can be used to map the location of the GEF32529 genes on a chromosome. The mapping of the GEF32529 sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, GEF32529 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the GEF32529 nucleotide sequences. Computer analysis of the GEF32529 sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the GEF32529 sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. *et al.* (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the GEF32529 nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a GEF32529 sequence to its chromosome include *in situ* hybridization (described in Fan, Y. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.*, Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding

sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. *et al.* (1987) *Nature,* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the GEF32529 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The GEF32529 sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the GEF32529 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the

present invention can be used to obtain such identification sequences from individuals and from tissue. The GEF32529 nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from GEF32529 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of GEF32529 Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, *e.g.*, hair or skin, or body fluids, *e.g.*, blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, *e.g.*, PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (*i.e.* another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique.

Examples of polynucleotide reagents include the GEF32529 nucleotide sequences or portions thereof, *e.g.*, fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases.

The GEF32529 nucleotide sequences described herein can further be used to provide polynucleotide reagents, *e.g.*, labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, *e.g.*, brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such GEF32529 probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, *e.g.*, GEF32529 primers or probes can be used to screen tissue culture for contamination (*i.e.* screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining GEF32529 polypeptide and/or nucleic acid expression as well as GEF32529 activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted GEF32529 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with GEF32529 polypeptide, nucleic acid expression or activity. For example, mutations in a GEF32529 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with GEF32529 polypeptide, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of GEF32529 in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of GEF32529 polypeptide or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting GEF32529 polypeptide or nucleic acid (*e.g.*, mRNA, or

genomic DNA) that encodes GEF32529 polypeptide such that the presence of GEF32529 polypeptide or nucleic acid is detected in the biological sample. In another aspect, the present invention provides a method for detecting the presence of GEF32529 activity in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of GEF32529 activity such that the presence of GEF32529 activity is detected in the biological sample. A preferred agent for detecting GEF32529 mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to GEF32529 mRNA or genomic DNA. The nucleic acid probe can be, for example, the GEF32529 nucleic acid set forth in SEQ ID NO:1 or 3, or the DNA insert of the plasmid deposited with ATCC as Accession Number _____, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to GEF32529 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting GEF32529 polypeptide is an antibody capable of binding to GEF32529 polypeptide, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')2) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect GEF32529 mRNA, polypeptide, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of GEF32529 mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of GEF32529 polypeptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of GEF32529 genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of GEF32529 polypeptide include introducing into a subject a labeled anti-GEF32529 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a GEF32529 polypeptide; (ii) aberrant expression of a gene encoding a GEF32529 polypeptide; (iii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a GEF32529 polypeptide, wherein a wild-type form of the gene encodes a polypeptide with a GEF32529 activity. "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes, but is not limited to, expression at non-wild type levels (*e.g.*, over or under expression); a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed (*e.g.*, increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage); a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene (*e.g.*, a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus).

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting GEF32529 polypeptide, mRNA, or genomic DNA, such that the presence of GEF32529 polypeptide, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of GEF32529 polypeptide, mRNA or genomic DNA in the control sample with the presence of GEF32529 polypeptide, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of GEF32529 in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting GEF32529 polypeptide or mRNA in a biological sample; means for determining the amount of GEF32529 in the sample; and means for comparing the amount of GEF32529 in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect GEF32529 polypeptide or nucleic acid.

2. Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant or unwanted GEF32529 expression or activity. As used herein, the term "aberrant" includes a GEF32529 expression or activity which deviates from the wild type GEF32529 expression or activity. Aberrant expression or activity includes increased or decreased expression or activity, as well as expression or activity which does not follow the wild type developmental pattern of expression or the subcellular pattern of expression. For example, aberrant GEF32529 expression or activity is intended to include the cases in which a mutation in the GEF32529 gene causes the GEF32529 gene to be under-expressed or over-expressed and situations in which such mutations result in a non-functional GEF32529 polypeptide or a polypeptide which does not function in a wild-type fashion, *e.g.*, a polypeptide which does not interact with a GEF32529 substrate, *e.g.*, a non-GEF subunit or ligand, or one which interacts with a non-GEF32529 substrate, *e.g.* a non-GEF subunit or ligand. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response, such as cellular proliferation. For example, the term unwanted includes a GEF32529 expression or activity which is undesirable in a subject.

The assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with a misregulation in GEF32529 polypeptide activity or nucleic acid expression, such as a GDP dissociation disorder (*e.g.*, a cell signaling, tumor inhibition, cytoskeletal organization, or cellular trafficking disorder). Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disorder associated with a misregulation in GEF32529 polypeptide activity or nucleic acid expression, such as a GDP dissociation disorder, or a cell signaling, tumor inhibition, cytoskeletal organization, or cellular trafficking disorder. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant or unwanted GEF32529 expression or activity in which a test sample is obtained from a subject and GEF32529 polypeptide or nucleic acid (*e.g.*, mRNA or genomic DNA) is detected, wherein the presence of GEF32529 polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted GEF32529 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted GEF32529 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a GDP dissociation disorder, or a cell signaling, tumor inhibition, cytoskeletal organization, or cellular trafficking disorder. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant or unwanted GEF32529 expression or activity in which a test sample is obtained and GEF32529 polypeptide or nucleic acid expression or activity is detected (*e.g.*, wherein the abundance of GEF32529 polypeptide or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant or unwanted GEF32529 expression or activity).

The methods of the invention can also be used to detect genetic alterations in a GEF32529 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in GEF32529 polypeptide activity or nucleic acid expression, such as a GDP dissociation disorder, or a cell signaling, tumor inhibition, cytoskeletal organization, or cellular trafficking disorder. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a GEF32529 -polypeptide, or the mis-expression of the GEF32529 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a GEF32529 gene; 2) an addition of one or more nucleotides to a GEF32529 gene; 3) a substitution of one or more nucleotides of a GEF32529 gene, 4) a chromosomal rearrangement of a GEF32529 gene; 5) an alteration in the level of a messenger RNA transcript of a GEF32529 gene, 6) aberrant modification of a GEF32529 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a GEF32529 gene, 8) a non-wild type level of a GEF32529-polypeptide, 9) allelic loss of a GEF32529 gene, and 10) inappropriate post-translational modification of a GEF32529-polypeptide. As described herein, there are a large number of assays known in the art which can be used for detecting alterations in a GEF32529 gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, *e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, *e.g.*, Landegran *et al.* (1988) *Science* 241:1077-1080; and Nakazawa *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the GEF32529-gene (see Abravaya *et al.* (1995) *Nucleic Acids Res* .23:675-682). This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a GEF32529 gene under conditions such that hybridization and amplification of the GEF32529-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. *et al.*, (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. *et al.* (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a GEF32529 gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in GEF32529 can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. *et al.* (1996) *Human Mutation* 7: 244-255; Kozal, M.J. *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in GEF32529 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T.

et al. supra. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the GEF32529 gene and detect mutations by comparing the sequence of the sample GEF32529 with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen *et al.* (1996) *Adv. Chromatogr.* 36:127-162; and Griffin *et al.* (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the GEF32529 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type GEF32529 sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in GEF32529 cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a GEF32529 sequence, *e.g.*, a wild-type GEF32529 sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in GEF32529 genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control GEF32529 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions

which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a GEF32529 gene.

Furthermore, any cell type or tissue in which GEF32529 is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs) on the expression or activity of a GEF32529 polypeptide (*e.g.*, the modulation of membrane excitability) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase GEF32529 gene expression, polypeptide levels, or upregulate GEF32529 activity, can be monitored in clinical trials of subjects exhibiting decreased GEF32529 gene expression, polypeptide levels, or downregulated GEF32529 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease GEF32529 gene expression, polypeptide levels, or downregulate GEF32529 activity, can be monitored in clinical trials of subjects exhibiting increased GEF32529 gene expression, polypeptide

levels, or upregulated GEF32529 activity. In such clinical trials, the expression or activity of a GEF32529 gene, and preferably, other genes that have been implicated in, for example, a GEF32529-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including GEF32529, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates GEF32529 activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on GEF32529-associated disorders (*e.g.*, disorders characterized by deregulated GEF activity), for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of GEF32529 and other genes implicated in the GEF32529-associated disorder, respectively. The levels of gene expression (*e.g.*, a gene expression pattern) can be quantified by northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of polypeptide produced, by one of the methods as described herein, or by measuring the levels of activity of GEF32529 or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) including the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a GEF32529 polypeptide, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the GEF32529 polypeptide, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the GEF32529 polypeptide, mRNA, or genomic DNA in the pre-administration sample with the GEF32529 polypeptide, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of GEF32529 to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of GEF32529 to lower levels than detected, *i.e.* to decrease the effectiveness of the agent. According to such an embodiment, GEF32529

expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

D. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted GEF32529 expression or activity, *e.g.* a GEF associated or GEF related disorder, for example, a cell signaling, tumor inhibition, cytoskeletal organization, or cellular trafficking disorder. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (*e.g.*, a patient's "drug response phenotype", or "drug response genotype"). Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the GEF32529 molecules of the present invention or GEF32529 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

Treatment is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease.

A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted GEF32529 expression or activity, by administering to the subject a GEF32529 or an agent which modulates GEF32529 expression or at least one GEF32529 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted GEF32529 expression or activity can be identified by, for example, any or a combination of diagnostic or

prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the GEF32529 aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of GEF32529 aberrancy, for example, a GEF32529, GEF32529 agonist or GEF32529 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating GEF32529 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell capable of expressing GEF32529 with an agent that modulates one or more of the activities of GEF32529 polypeptide activity associated with the cell, such that GEF32529 activity in the cell is modulated. An agent that modulates GEF32529 polypeptide activity can be an agent as described herein, such as a nucleic acid or a polypeptide, a naturally-occurring target molecule of a GEF32529 polypeptide (*e.g.*, a GEF32529 substrate), a GEF32529 antibody, a GEF32529 agonist or antagonist, a peptidomimetic of a GEF32529 agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more GEF32529 activities. Examples of such stimulatory agents include active GEF32529 polypeptide and a nucleic acid molecule encoding GEF32529 that has been introduced into the cell. In another embodiment, the agent inhibits one or more GEF32529 activities. Examples of such inhibitory agents include antisense GEF32529 nucleic acid molecules, anti-GEF32529 antibodies, and GEF32529 inhibitors. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a GEF32529 polypeptide or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, upregulates or downregulates) GEF32529 expression or activity. In another embodiment, the method involves administering a GEF32529 polypeptide or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted GEF32529 expression or activity.

Stimulation of GEF32529 activity is desirable in situations in which GEF32529 is abnormally downregulated and/or in which increased GEF32529 activity is likely to have a beneficial effect. Likewise, inhibition of GEF32529 activity is desirable in situations in which GEF32529 is abnormally upregulated and/or in which decreased GEF32529 activity is likely to have a beneficial effect.

3. Pharmacogenomics

The GEF32529 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on GEF32529 activity (*e.g.*, GEF32529 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) GEF32529-associated disorders (*e.g.*, proliferative disorders) associated with aberrant or unwanted GEF32529 activity. In conjunction with such treatment, pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a GEF32529 molecule or GEF32529 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a GEF32529 molecule or GEF32529 modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11): 983-985 and Linder, M.W. *et al.* (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (*e.g.*, a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known

single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (*e.g.*, a GEF32529 polypeptide of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (*e.g.*, a GEF32529 molecule or GEF32529 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a GEF32529 molecule or GEF32529 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

4. Use of GEF32529 Molecules as Surrogate Markers

The GEF32529 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the GEF32529 molecules of the invention may be detected, and may be correlated with one or more biological states *in vivo*. For example, the GEF32529 molecules of the invention may serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (*e.g.*, with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers may serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (*e.g.*, early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (*e.g.*, an assessment of cardiovascular disease may be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection may be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

The GEF32529 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a

pharmacodynamic marker may be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug may be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker may be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug may be sufficient to activate multiple rounds of marker (*e.g.*, a GEF32529 marker) transcription or expression, the amplified marker may be in a quantity which is more readily detectable than the drug itself. Also, the marker may be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-GEF32529 antibodies may be employed in an immune-based detection system for a GEF32529 polypeptide marker, or GEF32529-specific radiolabeled probes may be used to detect a GEF32529 mRNA marker. Furthermore, the use of a pharmacodynamic marker may offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am, J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

The GEF32529 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, *e.g.*, McLeod *et al.* (1999) *Eur. J. Cancer* 35(12): 1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, may be selected. For example, based on the presence or quantity of RNA, or polypeptide (*e.g.*, GEF32529 polypeptide or RNA) for specific tumor markers in a subject, a drug or course of treatment may be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in GEF32529 DNA may correlate GEF32529 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

5. Electronic Apparatus Readable Media and Arrays

Electronic apparatus readable media comprising GEF32529 sequence information is also provided. As used herein, "GEF32529 sequence information" refers to any nucleotide and/or amino acid sequence information particular to the GEF32529 molecules of the present invention, including but not limited to full-length nucleotide and/or amino acid sequences, partial nucleotide and/or amino acid sequences, polymorphic sequences including single nucleotide polymorphisms (SNPs), epitope sequences, and the like. Moreover, information "related to" said GEF32529 sequence information includes detection of the presence or absence of a sequence (*e.g.*, detection of expression of a sequence, fragment, polymorphism, etc.), determination of the level of a sequence (*e.g.*, detection of a level of expression, for example, a quantitative detection), detection of a reactivity to a sequence (*e.g.*, detection of protein expression and/or levels, for example, using a sequence-specific antibody), and the like. As used herein, "electronic apparatus readable media" refers to any suitable medium for storing, holding or containing data or information that can be read and accessed directly by an electronic apparatus. Such media can include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc; electronic storage media such as RAM, ROM, EPROM, EEPROM and the like; general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having recorded thereon GEF32529 sequence information of the present invention.

As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus or other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as a personal digital assistants (PDAs), cellular phone, pager and the like; and local and distributed processing systems.

As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the GEF32529 sequence information.

A variety of software programs and formats can be used to store the sequence information on the electronic apparatus readable medium. For example, the sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and MicroSoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2,

Sybase, Oracle, or the like, as well as in other forms. Any number of dataprocessor structuring formats (*e.g.*, text file or database) may be employed in order to obtain or create a medium having recorded thereon the GEF32529 sequence information.

By providing GEF32529 sequence information in readable form, one can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the sequence information in readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a GEF32529-associated disease or disorder or a pre-disposition to a GEF32529-associated disease or disorder, wherein the method comprises the steps of determining GEF32529 sequence information associated with the subject and based on the GEF32529 sequence information, determining whether the subject has a GEF32529-associated disease or disorder or a pre-disposition to a GEF32529-associated disease or disorder and/or recommending a particular treatment for the disease, disorder or pre-disease condition.

The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a GEF32529-associated disease or disorder or a pre-disposition to a disease associated with a GEF32529 wherein the method comprises the steps of determining GEF32529 sequence information associated with the subject, and based on the GEF32529 sequence information, determining whether the subject has a GEF32529-associated disease or disorder or a pre-disposition to a GEF32529-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

The present invention also provides in a network, a method for determining whether a subject has a GEF32529-associated disease or disorder or a pre-disposition to a GEF32529-associated disease or disorder associated with GEF32529, said method comprising the steps of receiving GEF32529 sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to GEF32529 and/or a GEF32529-associated disease or disorder, and based on one or more of the phenotypic information, the GEF32529 information (*e.g.*, sequence information and/or information related thereto), and the acquired information, determining whether the subject has a GEF32529-associated disease or disorder or a pre-disposition to a GEF32529-associated

disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The present invention also provides a business method for determining whether a subject has a GEF32529-associated disease or disorder or a pre-disposition to a GEF32529-associated disease or disorder, said method comprising the steps of receiving information related to GEF32529 (*e.g.*, sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to GEF32529 and/or related to a GEF32529-associated disease or disorder, and based on one or more of the phenotypic information, the GEF32529 information, and the acquired information, determining whether the subject has a GEF32529-associated disease or disorder or a pre-disposition to a GEF32529-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder or pre-disease condition.

The invention also includes an array comprising a GEF32529 sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be GEF32529. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

In addition to such qualitative determination, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression *per se* and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression between or among tissues. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a GEF32529-associated disease or disorder, progression of GEF32529-associated disease or disorder, and processes, such a cellular transformation associated with the GEF32529-associated disease or disorder.

The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (*e.g.*, ascertaining the effect of GEF32529 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (*e.g.*, including GEF32529) that could serve as a molecular target for diagnosis or therapeutic intervention.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, as well as the Figures and the Sequence Listing, are incorporated herein by reference.

EXAMPLES

EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMAN GEF32529 cDNA

In this example, the identification and characterization of the gene encoding human GEF32529 (clone 32529) is described.

Isolation of the human GEF32529 cDNA

The invention is based, at least in part, on the discovery of a human gene encoding a novel polypeptide, referred to herein as human GEF32529. The entire sequence of the human clone 32529 was determined and found to contain an open reading frame termed human "GEF32529." The nucleotide sequence of the human GEF32529 gene is set forth in Figure 1 and in the Sequence Listing as SEQ ID NO:1. The amino acid sequence of the human GEF32529 expression product is set forth in Figure 1 and in the Sequence Listing as SEQ ID NO: 2. The GEF32529 polypeptide comprises about 802 amino acids. The coding region (open reading frame) of SEQ ID NO:1 is set forth as SEQ ID NO:3. Clone 32529, comprising the coding region of human GEF32529, was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, and assigned Accession No. _____.

Analysis of the Human GEF32529 Molecules

A search using the polypeptide sequence of SEQ ID NO:2 was performed against the HMM database in PFAM (Figure 3) resulting in the identification of a GEF domain in the amino acid sequence of human GEF32529 at about residues 380-559 of SEQ ID NO:2 (score = 64.5), a potential PGAM domain in the amino acid sequence of human GEF32529 at about residues 592-598 of SEQ ID NO:2 (score = 5.2), a PH domain in the amino acid sequence of human GEF32529 at about residues 593-704 of SEQ ID NO:2 (score = 33.0), and a SH3 domain in the amino acid sequence of human GEF32529 at about residues 724-774 of SEQ ID NO:2 (score = 29.7).

A search using the polypeptide sequence of SEQ ID NO:2 was performed against the HMM database in SMART (Figure 3), a database of HMMs which has been revised and updated by Applicant, confirming the identification of the GEF domain in the amino acid sequence of human GEF32529 (*e.g.*, at about residues 380-559 of SEQ ID NO:2 (score = 158.4)), the PH domain in the amino acid sequence of human GEF32529 (*e.g.*, at about residues 593-706 of SEQ ID NO:2 (score = 32.9)), and the

SH3 domain in the amino acid sequence of human GEF32529 (*e.g.*, at about residues 718-775 of SEQ ID NO:2 (score = 47.6)).

The amino acid sequence of human GEF32529 was analyzed using the program PSORT (http://www.psort.nibb.ac.jp) to predict the localization of the proteins within the cell. This program assesses the presence of different targeting and localization amino acid sequences within the query sequence. The results of the analyses show that human GEF32529 may be localized to the nucleus or to the cytoplasm.

Searches of the amino acid sequence of human GEF32529 were further performed against the Prosite database. These searches resulted in the identification in the amino acid sequence of human GEF32529 of a potential N-glycosylation site, a potential glycosaminoglycan attachment site, a number of potential cAMP- and cGMP-dependent protein kinase phosphorylation sites, a number of potential protein kinase C phosphorylation sites, a number of potential casein kinase II phosphorylation sites, and a number of potential N-myristoylation sites.

A MEMSAT analysis of the polypeptide sequence of SEQ ID NO:2 was also performed, predicting a possible transmembrane domain in the amino acid sequence of human GEF32529 (SEQ ID NO:2) at about residues 390-406.

Further hits were identified by using the amino acid sequence of GEF32529 (SEQ ID NO:2) to search the ProDom database. Numerous matches against proteins and/or protein domains described as "TIM oncogene guanine nucleotide neuroblastoma factor exchange", "neuroblastoma", "KIAA0915", "BCDNA:GH03693 K07D4.7", "TIM guanine nucleotide oncogene factor exchange", "factor releasing guanine-nucleotide exchange proto-oncogene domain binding phorbol-ester", "polymerase subunit gamma III DNA", "receptor dopamine family polymorphism G-protein D4 D2C multigene coupled repeat", "Rho exchange CG1225 nucleotide factor guanine", "FRGA", "early immediate transcription factor response activated ETR101 growth inducible cycloheximide-induced", "CG10555", "transporter ABC", "QCCE-12673 brain cDNA", "membrane", "element transposable TN4556 transposon", "kinase serine/threonine serine/threonine-protein", "CG5606", "cell trophinin-associated repeat adhesion tastin trophinin-assisting", "UL71", "calcium binding", and the like were identified.

Tissue Distribution of human GEF32529 mRNA

This example describes the tissue distribution of human GEF32529 mRNA, as may be determined by *in situ* analysis using oligonucleotide probes based on the human GEF32529 sequence.

For *in situ* analysis, various tissues, *e.g.* tissues obtained from brain, are first frozen on dry ice. Ten-micrometer-thick sections of the tissues are postfixed with 4%

formaldehyde in DEPC treated 1X phosphate- buffered saline at room temperature for 10 minutes before being rinsed twice in DEPC 1X phosphate-buffered saline and once in 0.1 M triethanolamine-HCl (pH 8.0). Following incubation in 0.25% acetic anhydride-0.1 M triethanolamine-HCl for 10 minutes, sections are rinsed in DEPC 2X SSC (1X SSC is 0.15M NaCl plus 0.015M sodium citrate). Tissue is then dehydrated through a series of ethanol washes, incubated in 100% chloroform for 5 minutes, and then rinsed in 100% ethanol for 1 minute and 95% ethanol for 1 minute and allowed to air dry.

Hybridizations are performed with ^{35}S-radiolabeled (5 X 10^7 cpm/ml) cRNA probes. Probes are incubated in the presence of a solution containing 600 mM NaCl, 10 mM Tris (pH 7.5), 1 mM EDTA, 0.01% sheared salmon sperm DNA, 0.01% yeast tRNA, 0.05% yeast total RNA type X1, 1X Denhardt's solution, 50% formamide, 10% dextran sulfate, 100 mM dithiothreitol, 0.1% sodium dodecyl sulfate (SDS), and 0.1% sodium thiosulfate for 18 hours at 55°C.

After hybridization, slides are washed with 2X SSC. Sections are then sequentially incubated at 37°C in TNE (a solution containing 10 mM Tris-HCl (pH 7.6), 500 mM NaCl, and 1 mM EDTA), for 10 minutes, in TNE with 10µg of RNase A per ml for 30 minutes, and finally in TNE for 10 minutes. Slides are then rinsed with 2X SSC at room temperature, washed with 2X SSC at 50°C for 1 hour, washed with 0.2X SSC at 55°C for 1 hour, and 0.2X SSC at 60°C for 1 hour. Sections are then dehydrated rapidly through serial ethanol-0.3 M sodium acetate concentrations before being air dried and exposed to Kodak Biomax MR scientific imaging film for 24 hours and subsequently dipped in NB-2 photoemulsion and exposed at 4°C for 7 days before being developed and counter stained.

EXAMPLE 2: EXPRESSION OF RECOMBINANT GEF32529 POLYPEPTIDE IN BACTERIAL CELLS

In this example, human GEF32529 is expressed as a recombinant glutathione-S-transferase (GST) fusion polypeptide in *E. coli* and the fusion polypeptide is isolated and characterized. Specifically, GEF32529 is fused to GST and this fusion polypeptide is expressed in *E. coli*, *e.g.*, strain PEB199. Expression of the GST-GEF32529 fusion polypeptide in PEB199 is induced with IPTG. The recombinant fusion polypeptide is purified from crude bacterial lysates of the induced PEB199 strain by affinity chromatography on glutathione beads. Using polyacrylamide gel electrophoretic analysis of the polypeptide purified from the bacterial lysates, the molecular weight of the resultant fusion polypeptide is determined.

EXAMPLE 3: EXPRESSION OF RECOMBINANT GEF32529 POLYPEPTIDE IN COS CELLS

To express the human GEF32529 gene in COS cells, the pcDNA/Amp vector by Invitrogen Corporation (San Diego, CA) is used. This vector contains an SV40 origin of replication, an ampicillin resistance gene, an *E. coli* replication origin, a CMV promoter followed by a polylinker region, and an SV40 intron and polyadenylation site. A DNA fragment encoding the entire GEF32529 polypeptide and an HA tag (Wilson *et al.* (1984) *Cell* 37:767) or a FLAG tag fused in-frame to its 3' end of the fragment is cloned into the polylinker region of the vector, thereby placing the expression of the recombinant polypeptide under the control of the CMV promoter.

To construct the plasmid, the human GEF32529 DNA sequence is amplified by PCR using two primers. The 5' primer contains the restriction site of interest followed by approximately twenty nucleotides of the GEF32529 coding sequence starting from the initiation codon; the 3' end sequence contains complementary sequences to the other restriction site of interest, a translation stop codon, the HA tag or FLAG tag and the last 20 nucleotides of the GEF32529 coding sequence. The PCR amplified fragment and the pCDNA/Amp vector are digested with the appropriate restriction enzymes and the vector is dephosphorylated using the CIAP enzyme (New England Biolabs, Beverly, MA). Preferably the two restriction sites chosen are different so that the GEF32529 gene is inserted in the correct orientation. The ligation mixture is transformed into *E. coli* cells (strains HB101, DH5α, SURE, available from Stratagene Cloning Systems, La Jolla, CA, can be used), the transformed culture is plated on ampicillin media plates, and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment.

COS cells are subsequently transfected with the human GEF32529-pcDNA/Amp plasmid DNA using the calcium phosphate or calcium chloride co-precipitation methods, DEAE-dextran-mediated transfection, lipofection, or electroporation. Other suitable methods for transfecting host cells can be found in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989. The expression of the IC54420 polypeptide is detected by radiolabelling (^{35}S-methionine or ^{35}S-cysteine available from NEN, Boston, MA, can be used) and immunoprecipitation (Harlow, E. and Lane, D. *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988) using an HA specific monoclonal antibody. Briefly, the cells are labelled for 8 hours with ^{35}S-methionine (or ^{35}S-cysteine). The culture media are then collected and the cells are lysed using

detergents (RIPA buffer, 150 mM NaCl, 1% NP-40, 0.1% SDS, 0.5% DOC, 50 mM Tris, pH 7.5). Both the cell lysate and the culture media are precipitated with an HA specific monoclonal antibody. Precipitated polypeptides are then analyzed by SDS-PAGE.

Alternatively, DNA containing the human GEF32529 coding sequence is cloned directly into the polylinker of the pCDNA/Amp vector using the appropriate restriction sites. The resulting plasmid is transfected into COS cells in the manner described above, and the expression of the GEF32529 polypeptide is detected by radiolabelling and immunoprecipitation using a GEF32529-specific monoclonal antibody.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1; and

(b) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:3.

2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2.

3. An isolated nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____.

4. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2.

5. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to the nucleotide sequence of SEQ ID NO:1 or 3, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 30 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 60% identical to the amino acid sequence of SEQ ID NO:2; and

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 10 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2.

6. An isolated nucleic acid molecule which hybridizes to a complement of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.

7. An isolated nucleic acid molecule comprising a nucleotide sequence which is complementary to the nucleotide sequence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.

9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

10. The vector of claim 9, which is an expression vector.

11. A host cell transfected with the expression vector of claim 10.

12. A method of producing a polypeptide comprising culturing the host cell of claim 11 in an appropriate culture medium to, thereby, produce the polypeptide.

13. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 10 contiguous amino acids of SEQ ID NO:2;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a complement of a nucleic acid molecule consisting of SEQ ID NO:1 or 3 under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or 3; and

d) a polypeptide comprising an amino acid sequence which is at least 60% identical to the amino acid sequence of SEQ ID NO:2.

14. The isolated polypeptide of claim 13 comprising the amino acid sequence of SEQ ID NO:2.

15. The polypeptide of claim 13, further comprising heterologous amino acid sequences.

16. An antibody which selectively binds to a polypeptide of claim 13.

17. A method for detecting the presence of a polypeptide of claim 13 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 13 in the sample.

18. The method of claim 17, wherein the compound which binds to the polypeptide is an antibody.

19. A kit comprising a compound which selectively binds to a polypeptide of claim 13 and instructions for use.

20. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to a complement of the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to the complement of the nucleic acid molecule in the sample to thereby detect the presence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

21. The method of claim 20, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

22. A kit comprising a compound which selectively hybridizes to a complement of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

23. A method for identifying a compound which binds to a polypeptide of claim 13 comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

24. The method of claim 23, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detection of test compound/polypeptide binding;

b) detection of binding using a competition binding assay; and

c) detection of binding using an assay for GEF32529 activity.

25. A method for modulating the activity of a polypeptide of claim 13 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

26. A method for identifying a compound which modulates the activity of a polypeptide of claim 13 comprising:

a) contacting a polypeptide of claim 13 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

FIG. 1A

Input file Fbh32529FL.seq; Output File 32529.trans
Sequence length 3075

GCCCCACACAATACCCAGGAGCTTGCCTTGCTCGGCTCTGGGGCCATGCTGACATGCTGACATCGCCCCCTGAGGACTT

GGCTGCAACCCCAGAGCCCCCAGGGTGTCCCGGAGCCCTGGACCGTGCTGGCAGCTGGACGGAGCTCCCTGGCTGAGGG

```
                                M   D   C   G   P   P   A   T   L   Q   P   H   L     13
CCAGGTGGGTGGCAGAGCAAAAGAGGA ATG GAC TGT GGG CCA CCT GCT ACC CTC CAG CCC CAC CTG     39

 T   G   P   P   G   T   A   H   H   P   V   A   V   C   Q   Q   E   S   L   S     33
ACT GGG CCA CCT GGC ACT GCC CAC CAC CCT GTA GCA GTG TGC CAG CAG GAG AGT CTG TCC     99

 F   A   E   L   P   A   L   K   P   P   S   P   V   C   L   D   L   F   P   V     53
TTT GCA GAG CTG CCC GCC CTG AAG CCC CCG AGC CCA GTG TGT CTG GAC CTT TTC CCT GTT    159

 A   P   E   E   L   R   A   P   G   S   R   W   S   L   G   T   P   A   P   L     73
GCC CCA GAG GAG CTT CGG GCT CCT GGC AGC CGC TGG TCC CTG GGG ACC CCT GCC CCT CTC    219

 Q   G   L   L   W   P   L   S   P   G   G   S   D   T   E   I   T   S   G   G     93
CAA GGG TTG CTA TGG CCA TTA TCC CCA GGA GGC TCA GAT ACA GAG ATC ACC AGC GGG GGG    279

 M   R   P   S   R   A   G   S   W   P   H   C   P   G   A   Q   P   P   A   L    113
ATG CGG CCC AGC AGG GCT GGC AGC TGG CCA CAC TGT CCT GGT GCC CAG CCC CCA GCT CTG    339

 E   G   P   W   S   P   R   H   T   Q   P   Q   R   R   A   S   H   G   S   E    133
GAG GGA CCC TGG AGT CCC CGA CAC ACA CAG CCA CAG CGC CGG GCC AGC CAC GGC TCG GAG    399

 K   K   S   A   W   R   K   M   R   V   Y   Q   R   E   E   V   P   G   C   P    153
AAG AAG TCT GCC TGG CGC AAG ATG CGG GTG TAC CAG CGT GAA GAG GTC CCC GGC TGC CCC    459
```

FIG. 1B

```
 E   A   H   A   V   F   L   E   P   G   Q   V   V   Q   E   Q   A   L   S   T    173
GAG GCC CAC GCT GTC TTC CTA GAG CCT GGC CAG GTA GTG CAA GAG CAG GCC CTG AGC ACA    519

 E   E   P   R   V   E   L   S   G   S   T   R   V   S   L   E   G   P   E   R    193
GAG GAG CCC AGG GTG GAG TTG TCT GGG TCC ACC CGA GTG AGC CTC GAA GGT CCT GAG CGG    579

 R   R   F   S   A   S   E   L   M   T   R   L   H   S   S   L   R   L   G   R    213
AGG CGC TTC TCG GCA TCG GAG CTG ATG ACC CGG CTG CAC TCT TCT CTG CGC CTG GGG CGG    639

 N   S   A   A   R   A   L   I   S   G   S   G   T   G   A   A   R   E   G   K    233
AAT TCA GCA GCC CGG GCA CTC ATC TCT GGG TCA GGC ACC GGA GCA GCC CGG GAA GGG AAA    699

 A   S   G   M   E   A   R   S   V   E   M   S   G   D   R   V   S   R   P   A    253
GCA TCT GGA ATG GAG GCT CGA AGT GTA GAG ATG AGC GGG GAC CGG GTG TCG CGG CCA GCC    759

 P   G   D   S   R   E   G   D   W   S   E   P   R   L   D   T   Q   E   E   P    273
CCT GGT GAC TCA CGA GAG GGC GAT TGG TCC GAG CCC AGG CTA GAC ACA CAG GAA GAG CCG    819

 P   L   G   S   R   S   T   N   E   R   R   Q   S   R   F   L   L   N   S   V    293
CCT TTG GGG TCC AGG AGC ACC AAC GAG CGG CGC CAG TCT CGA TTC CTC CTT AAC TCC GTC    879

 L   Y   Q   E   Y   S   D   V   A   S   A   R   E   L   R   R   Q   Q   R   E    313
CTC TAT CAG GAA TAC AGC GAC GTG GCC AGC GCC CGC GAA CTG CGG CGG CAG CAG CGC GAG    939

 E   E   G   P   G   D   E   A   E   G   A   E   E   G   P   G   P   P   R   A    333
GAG GAG GGC CCG GGG GAC GAG GCC GAG GGC GCA GAG GAG GGG CCG GGG CCG CCG CGG GCC    999

 N   L   S   P   S   S   S   F   R   A   Q   R   S   A   R   G   S   T   F   S    353
AAC CTC TCC CCC AGC AGC TCC TTC CGG GCG CAG CGC TCG GCG CGA GGC TCC ACC TTC TCG   1059
```

FIG. 1D

```
 V   Q   E   C   N   A   S   V   Q   S   M   K   R   T   E   E   L   I   H   L    573
GTG CAG GAG TGC AAT GCT AGT GTA CAG TCC ATG AAG AGG ACA GAG GAA CTC ATC CAC CTG  1719

 S   K   K   I   H   F   E   G   K   I   F   P   L   I   S   Q   A   R   W   L    593
AGC AAG AAG ATC CAC TTT GAG GGC AAG ATT TTC CCG CTG ATC TCT CAG GCC CGC TGG CTG  1779

 V   R   H   G   E   L   V   E   L   A   P   L   P   A   A   P   P   A   K   L    613
GTT CGG CAT GGA GAG TTG GTA GAG CTG GCA CCA CTG CCT GCA GCA CCC CCT GCC AAG CTG  1839

 K   L   S   S   K   A   V   Y   L   H   L   F   N   D   C   L   L   L   S   R    633
AAG CTG TCC AGC AAG GCA GTC TAC CTC CAC CTC TTC AAT GAC TGC TTG CTG CTC TCT CGG  1899

 R   K   E   L   G   K   F   A   V   F   V   H   A   K   M   A   E   L   Q   V    653
CGG AAG GAG CTA GGG AAG TTT GCC GTT TTC GTC CAT GCC AAG ATG GCT GAG CTG CAG GTG  1959

 R   D   L   S   L   K   L   Q   G   I   P   G   H   V   F   L   L   Q   L   L    673
CGG GAC CTG AGC CTG AAG CTG CAG GGC ATC CCC GGC CAC GTG TTC CTC CTC CAG CTC CTC  2019

 H   G   Q   H   M   K   H   Q   F   L   L   R   A   R   T   E   S   E   K   Q    693
CAC GGG CAG CAC ATG AAG CAC CAG TTC CTG CTG CGG GCC CGG ACG GAA AGT GAG AAG CAG  2079

 R   W   I   S   A   L   C   P   S   S   P   Q   E   D   K   E   V   I   S   E    713
CGA TGG ATC TCA GCC TTG TGC CCC TCC AGC CCC CAG GAG GAC AAG GAG GTC ATC AGT GAG  2139

 G   E   D   C   P   Q   V   Q   C   V   R   T   Y   K   A   L   H   P   D   E    733
GGG GAA GAT TGC CCC CAG GTT CAG TGT GTT AGG ACA TAC AAG GCA CTG CAC CCA GAT GAG  2199

 L   T   L   E   K   T   D   I   L   S   V   R   T   W   T   S   D   G   W   L    753
CTG ACC TTG GAG AAG ACT GAC ATC CTG TCA GTG AGG ACC TGG ACC AGT GAC GGC TGG CTG  2259
```

FIG. 1C

```
L  W  Q  D  I  P  D  V  R  G  S  G  V  L  A  T  L  S  L  R   373
CTG TGG CAG GAT ATC CCC GAC GTA CGC GGC AGC GGC GTC CTG GCC ACG CTG AGC CTG CGG 1119

D  C  K  L  Q  E  A  K  F  E  L  I  T  S  E  A  S  Y  I  H   393
GAC TGC AAG CTG CAG GAG GCC AAG TTT GAG CTG ATC ACC TCC GAG GCC TCC TAC ATC CAC 1179

S  L  S  V  A  V  G  H  F  L  G  S  A  E  L  S  E  C  L  G   413
AGC CTG TCG GTG GCT GTG GGC CAC TTC TTA GGC TCT GCC GAG CTG AGC GAG TGT CTG GGG 1239

A  Q  D  K  Q  W  L  F  S  K  L  P  E  V  K  S  T  S  E  R   433
GCG CAG GAC AAG CAG TGG CTG TTT TCC AAA CTG CCC GAG GTC AAG AGC ACC AGC GAG AGG 1299

F  L  Q  D  L  E  Q  R  L  E  A  D  V  L  R  F  S  V  C  D   453
TTC CTG CAG GAC CTG GAG CAG CGG CTG GAG GCA GAT GTG CTG CGC TTC AGC GTG TGC GAC 1359

V  V  L  D  H  C  L  A  F  R  R  V  Y  L  P  Y  V  T  N  Q   473
GTG GTG CTG GAC CAC TGC CTG GCC TTC CGC AGA GTC TAC CTG CCC TAT GTC ACC AAC CAG 1419

A  Y  Q  E  R  T  Y  Q  R  L  L  L  E  N  P  R  F  P  G  I   493
GCC TAC CAG GAG CGC ACC TAC CAG CGC CTG CTC CTG GAG AAC CCC AGG TTC CCT GGC ATC 1479

L  A  R  L  E  E  S  P  V  C  Q  R  L  P  L  T  S  F  L  I   513
CTG GCT CGC CTG GAG GAG TCT CCT GTG TGC CAG CGT CTG CCC CTT ACC TCC TTC CTT ATC 1539

L  P  F  Q  R  I  T  R  L  K  M  L  V  E  N  I  L  K  R  T   533
CTG CCC TTC CAG AGG ATC ACC CGC CTC AAG ATG TTG GTG GAG AAC ATC CTG AAG CGG ACA 1599

A  Q  G  S  E  D  E  D  M  A  T  K  A  F  N  A  L  K  E  L   553
GCA CAG GGC TCT GAA GAC GAA GAC ATG GCC ACC AAG GCC TTC AAT GCG CTC AAG GAG CTG 1659
```

FIG. 1E

```
 E  G  V  R  L  A  D  G  E  K  G  W  V  P  Q  A  Y  V  E  E    773
GAA GGG GTC CGC CTG GCA GAT GGT GAG AAG GGG TGG GTG CCC CAG GCC TAT GTG GAA GAG  2319

 I  S  S  L  S  A  R  L  R  N  L  R  E  N  K  R  V  T  S  A    793
ATC AGC AGC CTC AGC GCC CGC CTC CGA AAC CTC CGG GAG AAT AAG CGA GTC ACA AGT GCC  2379

 T  S  K  L  G  E  A  P  V  *                                   803
ACC AGC AAA CTG GGG GAG GCT CCT GTG TGA                         2409

TGGGCAGCCATGGCCTAGGACCCCACCTCCATGCCTGGCTCCTGGATGGTCCTGCAGGGGCCTGCAGTGTCTCCATTCC

CCAAGCTGCTCCTGCTGGCACTTCGCTTCTGTGGCCTTGGCATTGAGGGCACAGGCTGGACACAGGAATGGGGGCGCCT

CCAGAGGGTCTCTCCGTCCTCATGCTCCTCAGTGTCCACACTTCAAGGCCAAGGATAGTTTCTTCCTCTGACATGGGA

CCATAACAGGTGATCACTGATACCTGGCAAAGACTGGGGCCCTCTCCTTCTATGTCCTCAATCCTGCCTGACTCTTGG

TCCTTCTGGCAGGGACCTGGCTGGGGAACGTTCTGGTGCTGATGGTGCTGGGCCCTATATGTATTTATATTATAGATCTG

GGGTGGGGTCTACCACGTCCAGTGGTCAAGGCCTCATTGGGTGTTGGTTGGTGTATGGTCTGTAAAGAGAATCCGAT

GATGCCT
```

FIG. 2

>32529
MDCGPPATLQPHLTGPPGTAHHPVAVCQQESLSFAELPALKPPSPVCLDLFPVAPEELRA
PGSRWSLGTPAPLQGLLWPLSPGGSDTEITSGGMRPSRAGSWPHCPGAQPPALEGRWSPR
HTQPQRRASHGSEKKSAWRKMRVYQREEVPGCPEAHAVFLEPGQVVQEQALSTEEPRVEL
SGSTRVSLEGPERRRFSASELMTRLHSSLRLGRNSAARALISGSGTGAAREGKASGMEAR
SVEMSGDRVSRPAPGDSREGDWSEPRLDTQEEPPLGSRSTNERRQSRFLLNSVLYQEYSD
VASARELRRQQREEEGPGDEAEGAEEGPGPPRANLSPSSSFRAQRSARGSTFSLWQDIPD
VRGSGVLATLSLRDCKLQEAKFELITSEASYIHSLSVAVGHFLGSAELSECLGAQDKQWL
FSKLPEVKSTSERFLQDLEQRLEADVLRFSVCDVVLDHCLAFRRVYLPYVTNQAYQERTY
QRLLLENPRFPGILARLEESPVCQRLPLTSFLILPFQRITRLKMLVENILKRTAQGSEDE
DMATKAFNALKELVQECNASVQSMKRTEELIHLSKKIHFEGKIFPLISQARWLVRHGELV
ELAPLPAAPPAKLKLSSKAVYLHLFNDCLLLSRRKELGKFAVFVHAKMAELQVRDLSLKL
QGIPGHVFLLQLLHGQHMKHQFLLRARTESEKQRWISALCPSSPQEDKEVISEGEDCPQV
QCVRTYKALHPDELTLEKTDILSVRTWTSDGWLEGVRLADGEKGWVPQAYVEEISSLSAR
LRNLRENKRVTSATSKLGEAPV

FIG. 3A

```
Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
hmmpfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 /prod/ddm/seqanal/PFAM/pfam6.2/Pfam
Sequence file:            /prod/ddm/wspace/orfanal/oa-script.8728.seq
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
 Query:  32529

Scores for sequence family classification (score includes all domains):
Model    Description                                        Score    E-value  N
-------- -----------                                        -----    ------- ---
RhoGEF   RhoGEF domain                                       64.5    2.3e-15   1
PH       PH domain                                           33.0    7.1e-08   1
SH3      SH3 domain                                          29.7    4.9e-05   1
PGAM     Phosphoglycerate mutase family                       5.2        2.8   1

Parsed for domains:
Model    Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
-------- ------- ----- -----    ----- -----       -----  -------
RhoGEF     1/1     380   559 ..     1   207 []    64.5  2.3e-15
PGAM       1/1     592   598 ..     1     7 [.     5.2      2.8
PH         1/1     593   704 ..     1    85 []    33.0  7.1e-08
SH3        1/1     724   774 ..     1    57 []    29.7  4.9e-05
```

FIG. 3B

```
Alignments of top-scoring domains:
RhoGEF: domain 1 of 1, from 380 to 559: score 64.5, E = 2.3e-15
                    *->vlkElleTEkkYvrdLeildnvym..kpLreaaisskpvltpddiet
                       + +El+++E+ Y+++L   + +++++++L e        l+++d++
       32529    380    AKFELITSEASYIHSLSVAVGHFLgsAELSE-------CLGAQDKQW 419

                    iFsNiediyefhreFLkssLearisssqfedlDekkiepsaprlGdlFlk
                    +Fs++++  +  +FL+  Le+r+++           ++    + d++l+
       32529    420 LFSKLPEVKSTSERFLQ-DLEQRLEA-----------DVLRFSVCDVVLD 457

                    lkepflqvYgeYcsnkpyAqelleklrqaasnpqFaefldeveassntgA
                    + ++f  vY +Y++n+ y + +  +l+'    np F  +l ++e s+ +
       32529    458 HCLAFRRVYLPYVTNQAYQERTYQRLL--LENPRFPGILARLEESPVCQ- 504

                    kddavkltLqsLLlkPvqRilrYpLLLkeLLkhtpegedqpdredLkkal
                         +l L s+L+ P qRi+r   L+  +Lk t    +++d+ ++ ka
       32529    505 -----RLPLTSFLILPFQRITRLKMLVENILKRTAQ--GSEDEDMATKAF 547

                    dllqdlaksiNe<-*
                     +l++l +  N+
       32529    548 NALKELVQECNA    559

PGAM: domain 1 of 1, from 592 to 598: score 5.2, E = 2.8
                    *->vLvRHGE<-*
                       +LvRHGE
       32529    592    WLVRHGE    598
```

FIG. 3C

```
PH: domain 1 of 1, from 593 to 704: score 33.0, E = 7.1e-08
                   *->vikeGwLlkks...........kswkkRyfvLfnnvLlyykdsk...
                      ++++G+L  + + +  ++ + + s k +y+ Lfn++Ll+ + +  ++
         32529   593    LVRHGELVELAplpaappaklkLSSKAVYLHLFNDCLLLSRRKElgk 639

                   kkpkgsipLsgcqvek.pd........kncFeirt.dr...tlllqaese
                   +       + ++qv++ + + ++ +++ + +++ ++++ ++++ll+a +e
         32529   640 FAVFVHAKMAELQVRDlSLklqgipghVFLLQLLHgQHmkhQFLLRARTE 689

                   eerkeWvkaiqsair<-*
                    e+++W++a++  +
         32529   690 SEKQRWISALCPSSP    704

SH3: domain 1 of 1, from 724 to 774: score 29.7, E = 4.9e-05
                   *->pkvvALYDYeaeesdELsFkkGDvitvleksdd.WWkgrlkgtgGke
                      ++++AL      ++dEL+++k D++ v   + d+W +g ++ +g ++
         32529   724    RTYKAL------HPDELTLEKTDILSVRTWTSDgWLEGVRLADG-EK 763

                   GlvPsNYVepv<-*
                   G+vP  YVe++
         32529   764 GWVPQAYVEEI    774
```

SEQUENCE LISTING

<110> Millennium Pharmaceuticals, Inc., *et al*

<120> 32529, A NOVEL HUMAN GUANINE NUCLEOTIDE EXCHANGE FACTOR FAMILY MEMBER AND USESE THEREOF

<130> MNI-175PC

<140>
<141>

<150> 60/218,983
<151> 2000-07-14

<160> 3

<170> PatentIn Ver. 2.0

<210> 1
<211> 3075
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (186)..(2591)

<400> 1
gccccacaca atacccagga gcttgccttg ctcggctctg gggccatgct gacatgctga 60

catcgccccc tgaggacttg gctgcaaccc cagagccccc agggtgtccc ggagccctgg 120

accgtgctgg cagctggacg gagctccctg gctgagggcc aggtgggtgg cagagcaaaa 180

gagga atg gac tgt ggg cca cct gct acc ctc cag ccc cac ctg act ggg 230
Met Asp Cys Gly Pro Pro Ala Thr Leu Gln Pro His Leu Thr Gly
1 5 10 15

cca cct ggc act gcc cac cac cct gta gca gtg tgc cag cag gag agt 278
Pro Pro Gly Thr Ala His His Pro Val Ala Val Cys Gln Gln Glu Ser
20 25 30

ctg tcc ttt gca gag ctg ccc gcc ctg aag ccc ccg agc cca gtg tgt 326
Leu Ser Phe Ala Glu Leu Pro Ala Leu Lys Pro Pro Ser Pro Val Cys
35 40 45

ctg gac ctt ttc cct gtt gcc cca gag gag ctt cgg gct cct ggc agc 374
Leu Asp Leu Phe Pro Val Ala Pro Glu Glu Leu Arg Ala Pro Gly Ser
50 55 60

cgc tgg tcc ctg ggg acc cct gcc cct ctc caa ggg ttg cta tgg cca 422
Arg Trp Ser Leu Gly Thr Pro Ala Pro Leu Gln Gly Leu Leu Trp Pro
65 70 75

tta tcc cca gga ggc tca gat aca gag atc acc agc ggg ggg atg cgg 470
Leu Ser Pro Gly Gly Ser Asp Thr Glu Ile Thr Ser Gly Gly Met Arg
80 85 90 95

```
ccc agc agg gct ggc agc tgg cca cac tgt cct ggt gcc cag ccc cca   518
Pro Ser Arg Ala Gly Ser Trp Pro His Cys Pro Gly Ala Gln Pro Pro
                100                 105                 110

gct ctg gag gga ccc tgg agt ccc cga cac aca cag cca cag cgc cgg   566
Ala Leu Glu Gly Pro Trp Ser Pro Arg His Thr Gln Pro Gln Arg Arg
            115                 120                 125

gcc agc cac ggc tcg gag aag aag tct gcc tgg cgc aag atg cgg gtg   614
Ala Ser His Gly Ser Glu Lys Lys Ser Ala Trp Arg Lys Met Arg Val
        130                 135                 140

tac cag cgt gaa gag gtc ccc ggc tgc ccc gag gcc cac gct gtc ttc   662
Tyr Gln Arg Glu Glu Val Pro Gly Cys Pro Glu Ala His Ala Val Phe
    145                 150                 155

cta gag cct ggc cag gta gtg caa gag cag gcc ctg agc aca gag gag   710
Leu Glu Pro Gly Gln Val Val Gln Glu Gln Ala Leu Ser Thr Glu Glu
160                 165                 170                 175

ccc agg gtg gag ttg tct ggg tcc acc cga gtg agc ctc gaa ggt cct   758
Pro Arg Val Glu Leu Ser Gly Ser Thr Arg Val Ser Leu Glu Gly Pro
                180                 185                 190

gag cgg agg cgc ttc tcg gca tcg gag ctg atg acc cgg ctg cac tct   806
Glu Arg Arg Arg Phe Ser Ala Ser Glu Leu Met Thr Arg Leu His Ser
            195                 200                 205

tct ctg cgc ctg ggg cgg aat tca gca gcc cgg gca ctc atc tct ggg   854
Ser Leu Arg Leu Gly Arg Asn Ser Ala Ala Arg Ala Leu Ile Ser Gly
        210                 215                 220

tca ggc acc gga gca gcc cgg gaa ggg aaa gca tct gga atg gag gct   902
Ser Gly Thr Gly Ala Ala Arg Glu Gly Lys Ala Ser Gly Met Glu Ala
    225                 230                 235

cga agt gta gag atg agc ggg gac cgg gtg tcg cgg cca gcc cct ggt   950
Arg Ser Val Glu Met Ser Gly Asp Arg Val Ser Arg Pro Ala Pro Gly
240                 245                 250                 255

gac tca cga gag ggc gat tgg tcc gag ccc agg cta gac aca cag gaa   998
Asp Ser Arg Glu Gly Asp Trp Ser Glu Pro Arg Leu Asp Thr Gln Glu
                260                 265                 270

gag ccg cct ttg ggg tcc agg agc acc aac gag cgg cgc cag tct cga  1046
Glu Pro Pro Leu Gly Ser Arg Ser Thr Asn Glu Arg Arg Gln Ser Arg
            275                 280                 285

ttc ctc ctt aac tcc gtc ctc tat cag gaa tac agc gac gtg gcc agc  1094
Phe Leu Leu Asn Ser Val Leu Tyr Gln Glu Tyr Ser Asp Val Ala Ser
        290                 295                 300

gcc cgc gaa ctg cgg cgg cag cag cgc gag gag gag ggc ccg ggg gac  1142
Ala Arg Glu Leu Arg Arg Gln Gln Arg Glu Glu Glu Gly Pro Gly Asp
    305                 310                 315

gag gcc gag ggc gca gag gag ggg ccg ggg ccg ccg cgg gcc aac ctc  1190
Glu Ala Glu Gly Ala Glu Glu Gly Pro Gly Pro Pro Arg Ala Asn Leu
320                 325                 330                 335
```

```
tcc ccc agc agc tcc ttc cgg gcg cag cgc tcg gcg cga ggc tcc acc   1238
Ser Pro Ser Ser Ser Phe Arg Ala Gln Arg Ser Ala Arg Gly Ser Thr
                340                 345                 350

ttc tcg ctg tgg cag gat atc ccc gac gta cgc ggc agc ggc gtc ctg   1286
Phe Ser Leu Trp Gln Asp Ile Pro Asp Val Arg Gly Ser Gly Val Leu
            355                 360                 365

gcc acg ctg agc ctg cgg gac tgc aag ctg cag gag gcc aag ttt gag   1334
Ala Thr Leu Ser Leu Arg Asp Cys Lys Leu Gln Glu Ala Lys Phe Glu
        370                 375                 380

ctg atc acc tcc gag gcc tcc tac atc cac agc ctg tcg gtg gct gtg   1382
Leu Ile Thr Ser Glu Ala Ser Tyr Ile His Ser Leu Ser Val Ala Val
    385                 390                 395

ggc cac ttc tta ggc tct gcc gag ctg agc gag tgt ctg ggg gcg cag   1430
Gly His Phe Leu Gly Ser Ala Glu Leu Ser Glu Cys Leu Gly Ala Gln
400                 405                 410                 415

gac aag cag tgg ctg ttt tcc aaa ctg ccc gag gtc aag agc acc agc   1478
Asp Lys Gln Trp Leu Phe Ser Lys Leu Pro Glu Val Lys Ser Thr Ser
                420                 425                 430

gag agg ttc ctg cag gac ctg gag cag cgg ctg gag gca gat gtg ctg   1526
Glu Arg Phe Leu Gln Asp Leu Glu Gln Arg Leu Glu Ala Asp Val Leu
            435                 440                 445

cgc ttc agc gtg tgc gac gtg gtg ctg gac cac tgc ctg gcc ttc cgc   1574
Arg Phe Ser Val Cys Asp Val Val Leu Asp His Cys Leu Ala Phe Arg
        450                 455                 460

aga gtc tac ctg ccc tat gtc acc aac cag gcc tac cag gag cgc acc   1622
Arg Val Tyr Leu Pro Tyr Val Thr Asn Gln Ala Tyr Gln Glu Arg Thr
    465                 470                 475

tac cag cgc ctg ctc ctg gag aac ccc agg ttc cct ggc atc ctg gct   1670
Tyr Gln Arg Leu Leu Leu Glu Asn Pro Arg Phe Pro Gly Ile Leu Ala
480                 485                 490                 495

cgc ctg gag gag tct cct gtg tgc cag cgt ctg ccc ctt acc tcc ttc   1718
Arg Leu Glu Glu Ser Pro Val Cys Gln Arg Leu Pro Leu Thr Ser Phe
                500                 505                 510

ctt atc ctg ccc ttc cag agg atc acc cgc ctc aag atg ttg gtg gag   1766
Leu Ile Leu Pro Phe Gln Arg Ile Thr Arg Leu Lys Met Leu Val Glu
            515                 520                 525

aac atc ctg aag cgg aca gca cag ggc tct gaa gac gaa gac atg gcc   1814
Asn Ile Leu Lys Arg Thr Ala Gln Gly Ser Glu Asp Glu Asp Met Ala
        530                 535                 540

acc aag gcc ttc aat gcg ctc aag gag ctg gtg cag gag tgc aat gct   1862
Thr Lys Ala Phe Asn Ala Leu Lys Glu Leu Val Gln Glu Cys Asn Ala
    545                 550                 555

agt gta cag tcc atg aag agg aca gag gaa ctc atc cac ctg agc aag   1910
Ser Val Gln Ser Met Lys Arg Thr Glu Glu Leu Ile His Leu Ser Lys
560                 565                 570                 575
```

```
aag atc cac ttt gag ggc aag att ttc ccg ctg atc tct cag gcc cgc   1958
Lys Ile His Phe Glu Gly Lys Ile Phe Pro Leu Ile Ser Gln Ala Arg
                580                 585                 590

tgg ctg gtt cgg cat gga gag ttg gta gag ctg gca cca ctg cct gca   2006
Trp Leu Val Arg His Gly Glu Leu Val Glu Leu Ala Pro Leu Pro Ala
            595                 600                 605

gca ccc cct gcc aag ctg aag ctg tcc agc aag gca gtc tac ctc cac   2054
Ala Pro Pro Ala Lys Leu Lys Leu Ser Ser Lys Ala Val Tyr Leu His
        610                 615                 620

gtc ttc aat gac tgc ttg ctg ctc tct cgg cgg aag gag cta ggg aag   2102
Leu Phe Asn Asp Cys Leu Leu Leu Ser Arg Arg Lys Glu Leu Gly Lys
    625                 630                 635

ttt gcc gtt ttc gtc cat gcc aag atg gct gag ctg cag gtg cgg gac   2150
Phe Ala Val Phe Val His Ala Lys Met Ala Glu Leu Gln Val Arg Asp
640                 645                 650                 655

ctg agc ctg aag ctg cag ggc atc ccc ggc cac gtg ttc ctc ctc cag   2198
Leu Ser Leu Lys Leu Gln Gly Ile Pro Gly His Val Phe Leu Leu Gln
                660                 665                 670

ctc ctc cac ggg cag cac atg aag cac cag ttc ctg ctg cgg gcc cgg   2246
Leu Leu His Gly Gln His Met Lys His Gln Phe Leu Leu Arg Ala Arg
            675                 680                 685

acg gaa agt gag aag cag cga tgg atc tca gcc ttg tgc ccc tcc agc   2294
Thr Glu Ser Glu Lys Gln Arg Trp Ile Ser Ala Leu Cys Pro Ser Ser
        690                 695                 700

ccc cag gag gac aag gag gtc atc agt gag ggg gaa gat tgc ccc cag   2342
Pro Gln Glu Asp Lys Glu Val Ile Ser Glu Gly Glu Asp Cys Pro Gln
    705                 710                 715

gtt cag tgt gtt agg aca tac aag gca ctg cac cca gat gag ctg acc   2390
Val Gln Cys Val Arg Thr Tyr Lys Ala Leu His Pro Asp Glu Leu Thr
720                 725                 730                 735

ttg gag aag act gac atc ctg tca gtg agg acc tgg acc agt gac ggc   2438
Leu Glu Lys Thr Asp Ile Leu Ser Val Arg Thr Trp Thr Ser Asp Gly
                740                 745                 750

tgg ctg gaa ggg gtc cgc ctg gca gat ggt gag aag ggg tgg gtg ccc   2486
Trp Leu Glu Gly Val Arg Leu Ala Asp Gly Glu Lys Gly Trp Val Pro
            755                 760                 765

cag gcc tat gtg gaa gag atc agc agc ctc agc gcc cgc ctc cga aac   2534
Gln Ala Tyr Val Glu Glu Ile Ser Ser Leu Ser Ala Arg Leu Arg Asn
        770                 775                 780

ctc cgg gag aat aag cga gtc aca agt gcc acc agc aaa ctg ggg gag   2582
Leu Arg Glu Asn Lys Arg Val Thr Ser Ala Thr Ser Lys Leu Gly Glu
    785                 790                 795

gct cct gtg tgatgggcag ccatggccta ggaccccacc tccatgcctg          2631
Ala Pro Val
800
```

gctcctggat ggtcctggag gggcctgcag tgtctccatt ccccaagctg ctcctgctgg 2691

cacttcgctt ctgtggcctt ggcattgagg gcacaggctg gacacaggaa tgggggcgcc 2751

tccagagggt ctctccgtcc tcatgctcct cagtgtccac acttcaaggc caaggatagt 2811

ttcttcctct gacatgggga ccataacagg tgatcactga tacctggcaa agactggggc 2871

cctctccttt ctatgtcctc aatcctgcct gactcttggt ccttctggca gggacctggc 2931

tggggaacgt tctggtgctg atggtgctgg gccctatatg tatatttata tagatctggg 2991

gtggggtcta gcacgtccag tggtcaaggc ctcattgggt gttggttggt gtgtatggtc 3051

tgtaaagaga atccgatgat gcct 3075

<210> 2
<211> 802
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Asp Cys Gly Pro Pro Ala Thr Leu Gln Pro His Leu Thr Gly Pro
  1               5                  10                  15

Pro Gly Thr Ala His His Pro Val Ala Val Cys Gln Gln Glu Ser Leu
             20                  25                  30

Ser Phe Ala Glu Leu Pro Ala Leu Lys Pro Pro Ser Pro Val Cys Leu
         35                  40                  45

Asp Leu Phe Pro Val Ala Pro Glu Glu Leu Arg Ala Pro Gly Ser Arg
     50                  55                  60

Trp Ser Leu Gly Thr Pro Ala Pro Leu Gln Gly Leu Leu Trp Pro Leu
 65                  70                  75                  80

Ser Pro Gly Gly Ser Asp Thr Glu Ile Thr Ser Gly Gly Met Arg Pro
                 85                  90                  95

Ser Arg Ala Gly Ser Trp Pro His Cys Pro Gly Ala Gln Pro Pro Ala
            100                 105                 110

Leu Glu Gly Pro Trp Ser Pro Arg His Thr Gln Pro Gln Arg Arg Ala
        115                 120                 125

Ser His Gly Ser Glu Lys Lys Ser Ala Trp Arg Lys Met Arg Val Tyr
    130                 135                 140

Gln Arg Glu Glu Val Pro Gly Cys Pro Glu Ala His Ala Val Phe Leu
145                 150                 155                 160

Glu Pro Gly Gln Val Val Gln Glu Gln Ala Leu Ser Thr Glu Glu Pro
                165                 170                 175

Arg Val Glu Leu Ser Gly Ser Thr Arg Val Ser Leu Glu Gly Pro Glu
            180                 185                 190

Arg Arg Arg Phe Ser Ala Ser Glu Leu Met Thr Arg Leu His Ser Ser
        195                 200                 205
```

```
Leu Arg Leu Gly Arg Asn Ser Ala Ala Arg Ala Leu Ile Ser Gly Ser
    210                 215                 220

Gly Thr Gly Ala Ala Arg Glu Gly Lys Ala Ser Gly Met Glu Ala Arg
225                 230                 235                 240

Ser Val Glu Met Ser Gly Asp Arg Val Ser Arg Pro Ala Pro Gly Asp
                245                 250                 255

Ser Arg Glu Gly Asp Trp Ser Glu Pro Arg Leu Asp Thr Gln Glu Glu
            260                 265                 270

Pro Pro Leu Gly Ser Arg Ser Thr Asn Glu Arg Arg Gln Ser Arg Phe
        275                 280                 285

Leu Leu Asn Ser Val Leu Tyr Gln Glu Tyr Ser Asp Val Ala Ser Ala
    290                 295                 300

Arg Glu Leu Arg Arg Gln Gln Arg Glu Glu Glu Gly Pro Gly Asp Glu
305                 310                 315                 320

Ala Glu Gly Ala Glu Glu Gly Pro Gly Pro Pro Arg Ala Asn Leu Ser
                325                 330                 335

Pro Ser Ser Ser Phe Arg Ala Gln Arg Ser Ala Arg Gly Ser Thr Phe
            340                 345                 350

Ser Leu Trp Gln Asp Ile Pro Asp Val Arg Gly Ser Gly Val Leu Ala
        355                 360                 365

Thr Leu Ser Leu Arg Asp Cys Lys Leu Gln Glu Ala Lys Phe Glu Leu
    370                 375                 380

Ile Thr Ser Glu Ala Ser Tyr Ile His Ser Leu Ser Val Ala Val Gly
385                 390                 395                 400

His Phe Leu Gly Ser Ala Glu Leu Ser Glu Cys Leu Gly Ala Gln Asp
                405                 410                 415

Lys Gln Trp Leu Phe Ser Lys Leu Pro Glu Val Lys Ser Thr Ser Glu
            420                 425                 430

Arg Phe Leu Gln Asp Leu Glu Gln Arg Leu Glu Ala Asp Val Leu Arg
        435                 440                 445

Phe Ser Val Cys Asp Val Val Leu Asp His Cys Leu Ala Phe Arg Arg
    450                 455                 460

Val Tyr Leu Pro Tyr Val Thr Asn Gln Ala Tyr Gln Glu Arg Thr Tyr
465                 470                 475                 480

Gln Arg Leu Leu Leu Glu Asn Pro Arg Phe Pro Gly Ile Leu Ala Arg
                485                 490                 495

Leu Glu Glu Ser Pro Val Cys Gln Arg Leu Pro Leu Thr Ser Phe Leu
            500                 505                 510

Ile Leu Pro Phe Gln Arg Ile Thr Arg Leu Lys Met Leu Val Glu Asn
        515                 520                 525
```

```
Ile Leu Lys Arg Thr Ala Gln Gly Ser Glu Asp Glu Asp Met Ala Thr
    530                 535                 540

Lys Ala Phe Asn Ala Leu Lys Glu Leu Val Gln Glu Cys Asn Ala Ser
545                 550                 555                 560

Val Gln Ser Met Lys Arg Thr Glu Glu Leu Ile His Leu Ser Lys Lys
                565                 570                 575

Ile His Phe Glu Gly Lys Ile Phe Pro Leu Ile Ser Gln Ala Arg Trp
            580                 585                 590

Leu Val Arg His Gly Glu Leu Val Glu Leu Ala Pro Leu Pro Ala Ala
        595                 600                 605

Pro Pro Ala Lys Leu Lys Leu Ser Ser Lys Ala Val Tyr Leu His Leu
    610                 615                 620

Phe Asn Asp Cys Leu Leu Leu Ser Arg Arg Lys Glu Leu Gly Lys Phe
625                 630                 635                 640

Ala Val Phe Val His Ala Lys Met Ala Glu Leu Gln Val Arg Asp Leu
                645                 650                 655

Ser Leu Lys Leu Gln Gly Ile Pro Gly His Val Phe Leu Leu Gln Leu
            660                 665                 670

Leu His Gly Gln His Met Lys His Gln Phe Leu Leu Arg Ala Arg Thr
        675                 680                 685

Glu Ser Glu Lys Gln Arg Trp Ile Ser Ala Leu Cys Pro Ser Ser Pro
    690                 695                 700

Gln Glu Asp Lys Glu Val Ile Ser Glu Gly Glu Asp Cys Pro Gln Val
705                 710                 715                 720

Gln Cys Val Arg Thr Tyr Lys Ala Leu His Pro Asp Glu Leu Thr Leu
                725                 730                 735

Glu Lys Thr Asp Ile Leu Ser Val Arg Thr Trp Thr Ser Asp Gly Trp
            740                 745                 750

Leu Glu Gly Val Arg Leu Ala Asp Gly Glu Lys Gly Trp Val Pro Gln
        755                 760                 765

Ala Tyr Val Glu Glu Ile Ser Ser Leu Ser Ala Arg Leu Arg Asn Leu
    770                 775                 780

Arg Glu Asn Lys Arg Val Thr Ser Ala Thr Ser Lys Leu Gly Glu Ala
785                 790                 795                 800

Pro Val
```

<210> 3
<211> 2406
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (1)..(2406)

<400> 3

```
atg gac tgt ggg cca cct gct acc ctc cag ccc cac ctg act ggg cca   48
Met Asp Cys Gly Pro Pro Ala Thr Leu Gln Pro His Leu Thr Gly Pro
  1               5                  10                  15

cct ggc act gcc cac cac cct gta gca gtg tgc cag cag gag agt ctg   96
Pro Gly Thr Ala His His Pro Val Ala Val Cys Gln Gln Glu Ser Leu
             20                  25                  30

tcc ttt gca gag ctg ccc gcc ctg aag ccc ccg agc cca gtg tgt ctg  144
Ser Phe Ala Glu Leu Pro Ala Leu Lys Pro Pro Ser Pro Val Cys Leu
         35                  40                  45

gac ctt ttc cct gtt gcc cca gag gag ctt cgg gct cct ggc agc cgc  192
Asp Leu Phe Pro Val Ala Pro Glu Glu Leu Arg Ala Pro Gly Ser Arg
     50                  55                  60

tgg tcc ctg ggg acc cct gcc cct ctc caa ggg ttg cta tgg cca tta  240
Trp Ser Leu Gly Thr Pro Ala Pro Leu Gln Gly Leu Leu Trp Pro Leu
 65                  70                  75                  80

tcc cca gga ggc tca gat aca gag atc acc agc ggg ggg atg cgg ccc  288
Ser Pro Gly Gly Ser Asp Thr Glu Ile Thr Ser Gly Gly Met Arg Pro
                 85                  90                  95

agc agg gct ggc agc tgg cca cac tgt cct ggt gcc cag ccc cca gct  336
Ser Arg Ala Gly Ser Trp Pro His Cys Pro Gly Ala Gln Pro Pro Ala
            100                 105                 110

ctg gag gga ccc tgg agt ccc cga cac aca cag cca cag cgc cgg gcc  384
Leu Glu Gly Pro Trp Ser Pro Arg His Thr Gln Pro Gln Arg Arg Ala
        115                 120                 125

agc cac ggc tcg gag aag aag tct gcc tgg cgc aag atg cgg gtg tac  432
Ser His Gly Ser Glu Lys Lys Ser Ala Trp Arg Lys Met Arg Val Tyr
    130                 135                 140

cag cgt gaa gag gtc ccc ggc tgc ccc gag gcc cac gct gtc ttc cta  480
Gln Arg Glu Glu Val Pro Gly Cys Pro Glu Ala His Ala Val Phe Leu
145                 150                 155                 160

gag cct ggc cag gta gtg caa gag cag gcc ctg agc aca gag gag ccc  528
Glu Pro Gly Gln Val Val Gln Glu Gln Ala Leu Ser Thr Glu Glu Pro
                165                 170                 175

agg gtg gag ttg tct ggg tcc acc cga gtg agc ctc gaa ggt cct gag  576
Arg Val Glu Leu Ser Gly Ser Thr Arg Val Ser Leu Glu Gly Pro Glu
            180                 185                 190

cgg agg cgc ttc tcg gca tcg gag ctg atg acc cgg ctg cac tct tct  624
Arg Arg Arg Phe Ser Ala Ser Glu Leu Met Thr Arg Leu His Ser Ser
        195                 200                 205

ctg cgc ctg ggg cgg aat tca gca gcc cgg gca ctc atc tct ggg tca  672
Leu Arg Leu Gly Arg Asn Ser Ala Ala Arg Ala Leu Ile Ser Gly Ser
    210                 215                 220
```

```
ggc acc gga gca gcc cgg gaa ggg aaa gca tct gga atg gag gct cga   720
Gly Thr Gly Ala Ala Arg Glu Gly Lys Ala Ser Gly Met Glu Ala Arg
225                 230                 235                 240

agt gta gag atg agc ggg gac cgg gtg tcg cgg cca gcc cct ggt gac   768
Ser Val Glu Met Ser Gly Asp Arg Val Ser Arg Pro Ala Pro Gly Asp
                245                 250                 255

tca cga gag ggc gat tgg tcc gag ccc agg cta gac aca cag gaa gag   816
Ser Arg Glu Gly Asp Trp Ser Glu Pro Arg Leu Asp Thr Gln Glu Glu
            260                 265                 270

ccg cct ttg ggg tcc agg agc acc aac gag cgg cgc cag tct cga ttc   864
Pro Pro Leu Gly Ser Arg Ser Thr Asn Glu Arg Arg Gln Ser Arg Phe
        275                 280                 285

ctc ctt aac tcc gtc ctc tat cag gaa tac agc gac gtg gcc agc gcc   912
Leu Leu Asn Ser Val Leu Tyr Gln Glu Tyr Ser Asp Val Ala Ser Ala
    290                 295                 300

cgc gaa ctg cgg cgg cag cag cgc gag gag gag ggc ccg ggg gac gag   960
Arg Glu Leu Arg Arg Gln Gln Arg Glu Glu Glu Gly Pro Gly Asp Glu
305                 310                 315                 320

gcc gag ggc gca gag gag ggg ccg ggg ccg ccg cgg gcc aac ctc tcc  1008
Ala Glu Gly Ala Glu Glu Gly Pro Gly Pro Pro Arg Ala Asn Leu Ser
                325                 330                 335

ccc agc agc tcc ttc cgg gcg cag cgc tcg gcg cga ggc tcc acc ttc  1056
Pro Ser Ser Ser Phe Arg Ala Gln Arg Ser Ala Arg Gly Ser Thr Phe
            340                 345                 350

tcg ctg tgg cag gat atc ccc gac gta cgc ggc agc ggc gtc ctg gcc  1104
Ser Leu Trp Gln Asp Ile Pro Asp Val Arg Gly Ser Gly Val Leu Ala
        355                 360                 365

acg ctg agc ctg cgg gac tgc aag ctg cag gag gcc aag ttt gag ctg  1152
Thr Leu Ser Leu Arg Asp Cys Lys Leu Gln Glu Ala Lys Phe Glu Leu
    370                 375                 380

atc acc tcc gag gcc tcc tac atc cac agc ctg tcg gtg gct gtg ggc  1200
Ile Thr Ser Glu Ala Ser Tyr Ile His Ser Leu Ser Val Ala Val Gly
385                 390                 395                 400

cac ttc tta ggc tct gcc gag ctg agc gag tgt ctg ggg gcg cag gac  1248
His Phe Leu Gly Ser Ala Glu Leu Ser Glu Cys Leu Gly Ala Gln Asp
                405                 410                 415

aag cag tgg ctg ttt tcc aaa ctg ccc gag gtc aag agc acc agc gag  1296
Lys Gln Trp Leu Phe Ser Lys Leu Pro Glu Val Lys Ser Thr Ser Glu
            420                 425                 430

agg ttc ctg cag gac ctg gag cag cgg ctg gag gca gat gtg ctg cgc  1344
Arg Phe Leu Gln Asp Leu Glu Gln Arg Leu Glu Ala Asp Val Leu Arg
        435                 440                 445

ttc agc gtg tgc gac gtg gtg ctg gac cac tgc ctg gcc ttc cgc aga  1392
Phe Ser Val Cys Asp Val Val Leu Asp His Cys Leu Ala Phe Arg Arg
    450                 455                 460
```

```
gtc tac ctg ccc tat gtc acc aac cag gcc tac cag gag cgc acc tac   1440
Val Tyr Leu Pro Tyr Val Thr Asn Gln Ala Tyr Gln Glu Arg Thr Tyr
465                 470                 475                 480

cag cgc ctg ctc ctg gag aac ccc agg ttc cct ggc atc ctg gct cgc   1488
Gln Arg Leu Leu Leu Glu Asn Pro Arg Phe Pro Gly Ile Leu Ala Arg
                485                 490                 495

ctg gag gag tct cct gtg tgc cag cgt ctg ccc ctt acc tcc ttc ctt   1536
Leu Glu Glu Ser Pro Val Cys Gln Arg Leu Pro Leu Thr Ser Phe Leu
            500                 505                 510

atc ctg ccg ttc cag agg atc acc agc ctc aag atg ttg gtg gag aac   1584
Ile Leu Pro Phe Gln Arg Ile Thr Arg Leu Lys Met Leu Val Glu Asn
        515                 520                 525

atc ctg aag cgg aca gca cag ggc tct gaa gac gaa gac atg gcc acc   1632
Ile Leu Lys Arg Thr Ala Gln Gly Ser Glu Asp Glu Asp Met Ala Thr
    530                 535                 540

aag gcc ttc aat gcg ctc aag gag ctg gtg cag gag tgc aat gct agt   1680
Lys Ala Phe Asn Ala Leu Lys Glu Leu Val Gln Glu Cys Asn Ala Ser
545                 550                 555                 560

gta cag tcc atg aag agg aca gag gaa ctc atc cac ctg agc aag aag   1728
Val Gln Ser Met Lys Arg Thr Glu Glu Leu Ile His Leu Ser Lys Lys
                565                 570                 575

atc cac ttt gag ggc aag att ttc ccg ctg atc tct cag gcc cgc tgg   1776
Ile His Phe Glu Gly Lys Ile Phe Pro Leu Ile Ser Gln Ala Arg Trp
            580                 585                 590

ctg gtt cgg cat gga gag ttg gta gag ctg gca cca ctg cct gca gca   1824
Leu Val Arg His Gly Glu Leu Val Glu Leu Ala Pro Leu Pro Ala Ala
        595                 600                 605

ccc cct gcc aag ctg aag ctg tcc agc aag gca gtc tac ctc cac ctc   1872
Pro Pro Ala Lys Leu Lys Leu Ser Ser Lys Ala Val Tyr Leu His Leu
    610                 615                 620

ttc aat gac tgc ttg ctg ctc tct cgg cgg aag gag cta ggg aag ttt   1920
Phe Asn Asp Cys Leu Leu Leu Ser Arg Arg Lys Glu Leu Gly Lys Phe
625                 630                 635                 640

gcc gtt ttc gtc cat gcc aag atg gct gag ctg cag gtg cgg gac ctg   1968
Ala Val Phe Val His Ala Lys Met Ala Glu Leu Gln Val Arg Asp Leu
                645                 650                 655

agc ctg aag ctg cag ggc atc ccc ggc cac gtg ttc ctc ctc cag ctc   2016
Ser Leu Lys Leu Gln Gly Ile Pro Gly His Val Phe Leu Leu Gln Leu
            660                 665                 670

ctc cac ggg cag cac atg aag cac cag ttc ctg ctg cgg gcc cgg acg   2064
Leu His Gly Gln His Met Lys His Gln Phe Leu Leu Arg Ala Arg Thr
        675                 680                 685

gaa agt gag aag cag cga tgg atc tca gcc ttg tgc ccc tcc agc ccc   2112
Glu Ser Glu Lys Gln Arg Trp Ile Ser Ala Leu Cys Pro Ser Ser Pro
    690                 695                 700
```

```
cag gag gac aag gag gtc atc agt gag ggg gaa gat tgc ccc cag gtt   2160
Gln Glu Asp Lys Glu Val Ile Ser Glu Gly Glu Asp Cys Pro Gln Val
705                 710                 715                 720

cag tgt gtt agg aca tac aag gca ctg cac cca gat gag ctg acc ttg   2208
Gln Cys Val Arg Thr Tyr Lys Ala Leu His Pro Asp Glu Leu Thr Leu
                725                 730                 735

gag aag act gac atc ctg tca gtg agg acc tgg acc agt gac ggc tgg   2256
Glu Lys Thr Asp Ile Leu Ser Val Arg Thr Trp Thr Ser Asp Gly Trp
            740                 745                 750

ctg gaa ggg gtc cgc ctg gca gat ggt gag aag ggg tgg gtg ccc cag   2304
Leu Glu Gly Val Arg Leu Ala Asp Gly Glu Lys Gly Trp Val Pro Gln
        755                 760                 765

gcc tat gtg gaa gag atc agc agc ctc agc gcc cgc ctc cga aac ctc   2352
Ala Tyr Val Glu Glu Ile Ser Ser Leu Ser Ala Arg Leu Arg Asn Leu
    770                 775                 780

cgg gag aat aag cga gtc aca agt gcc acc agc aaa ctg ggg gag gct   2400
Arg Glu Asn Lys Arg Val Thr Ser Ala Thr Ser Lys Leu Gly Glu Ala
785                 790                 795                 800

cct gtg                                                          2406
Pro Val
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 January 2002 (24.01.2002)

PCT

(10) International Publication Number
WO 02/06325 A3

(51) International Patent Classification[7]: C07K 14/47, 14/82, C12N 15/12, G01N 33/53, C12Q 1/68, C07K 16/18, 16/32

(21) International Application Number: PCT/US01/22290

(22) International Filing Date: 16 July 2001 (16.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/218,383 14 July 2000 (14.07.2000) US

(71) Applicant *(for all designated States except US)*: **MILLENNIUM PHARMACEUTICALS, INC.** [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **MEYERS, Rachel** [US/US]; 115 Devonshire Road, Newton, MA 02468 (US).

(74) Agents: **MANDRAGOURAS, Amy, E.** et al.; Lahive & Cockfield, LLP, 28 State Street, Boston, MA 02109 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
13 June 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/06325 A3

(54) Title: 32529, A HUMAN GUANINE NUCLEOTIDE EXCHANGE FACTOR FAMILY MEMBER AND USES THEREOF

(57) Abstract: The invention provides isolated nucleic acids molecules, designated GEF32529 nucleic acid molecules, which encode novel human guanine nucleotide exchange factor (GEF) molecules. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing GEF32529 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic animals in which a GEF32529 gene has been introduced or disrupted. The invention still further provides isolated GEF32529 polypeptides, fusion polypeptides, antigenic peptides and anti-GEF32529 antibodies. Diagnostic methods utilizing compositions of the invention are also provided.

INTERNATIONAL SEARCH REPORT

Inter ›nal Application No
PCT/US 01/22290

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/47 C07K14/82 C12N15/12 G01N33/53 C12Q1/68
C07K16/18 C07K16/32

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12N G01N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, EMBL, SEQUENCE SEARCH, BIOSIS, MEDLINE, PAJ, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! Accession AW675258, 12 April 2000 (2000-04-12) "ba61e11.y1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:2901068 5' similar to SW:TIM_HUMAN Q12774 PROBABLE GUANINE NUCLEOTIDE REGULATORY PROTEIN TIM ;, mRNA sequence." XP002194015 the whole document --- -/--	1-15

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
25 March 2002	09/04/2002

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Lejeune, R

2

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! Accession AL137736, 27 January 2000 (2000-01-27) OTTENWAELDER B ET AL: "Homo sapiens mRNA; cDNA DKFZp586P2321 (from clone DKFZp586P2321)" XP002194016 the whole document	5-8
X	CHAN A M-L ET AL: "EXPRESSION CDNA CLONING OF A NOVEL ONCOGENE WITH SEQUENCE SIMILARITY TO REGULATORS OF SMALL GTP-BINDING PROTEINS" ONCOGENE, BASINGSTOKE, HANTS, GB, vol. 9, no. 4, 1 April 1994 (1994-04-01), pages 1057-1068, XP002059291 ISSN: 0950-9232 the whole document	5-13
P,X	WO 01 46231 A (BURGESS CATHERINE E ;CURAGEN CORP (US)) 28 June 2001 (2001-06-28) page 8 -page 16	1-15
E	WO 02 06315 A (COMPUGEN LTD ;FREILICH SHIRI (IL); MINTZ LIAT (IL); BERNSTEIN JEAN) 24 January 2002 (2002-01-24) seq id 15 and 143 page 20	1-15

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 25 (partially)

Claim 25 refers to a compound which binds to the polypeptide seq id 2 without giving a true technical characterization. Moreover, no such compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject matter is not sufficiently disclosed and supported.
No search can be carried out for such speculative claims whose wording is, in fact a mere recitation of the results to be achieved.
A partial search has been carried out as far as the antagonist/agonist/inhibitor is an antibody to the polypeptide.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

...ormation on patent family members

Inter ...nal Application No
PCT/US 01/22290

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0146231 A	28-06-2001	AU 2287101 A	03-07-2001
		WO 0146231 A2	28-06-2001
WO 0206315 A	24-01-2002	WO 0206315 A2	24-01-2002